# When Children Chat with Machine Translated Text: Problems, Possibilities, Potential

Dev Raj Lamichhane
University of Central Lancashire
Preston, UK
drlamichhane1@uclan.ac.uk

Janet C. Read
University of Central Lancashire
Preston, UK
jcread@uclan.ac.uk

I. Scott MacKenzie
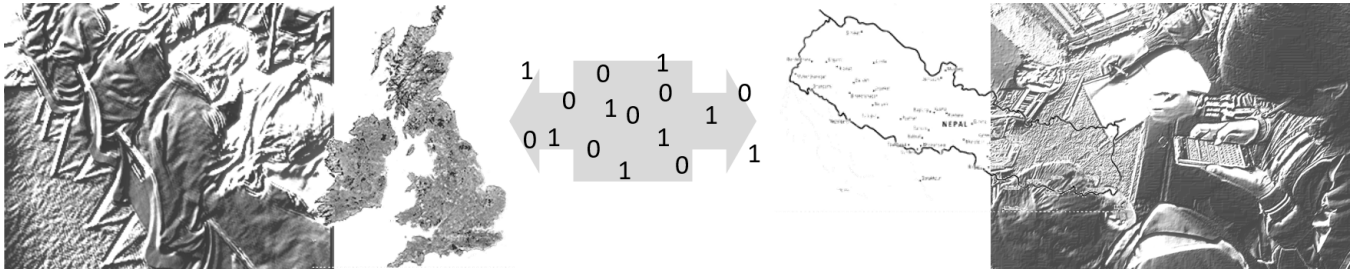York University
Toronto, Canada
mack@yorku.ca

**Figure 1: When Children Chat with Machine Translated Text.**

## ABSTRACT

Two cross-lingual (Nepalese and English) letter exchanges took place between school children from Nepal and England, using Digipal; an Android chatting application. Digipal uses Google Translate to enable children to read and reply in their native language. In two studies we analysed the errors made and the effect of errors on children's understanding and on the flow of conversation. We found that errors of input negatively affected translation, although this can be reduced through initial grammar cleaning. We highlight features of children's text that cause errors in translation whilst showing how children worked with and around these errors. Errors sometimes added humour and contributed to continuing the conversations.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *Usability testing*; *User studies*.

## KEYWORDS

Child Computer Interaction, Machine Translation, Text Entry, Errors

## 1 INTRODUCTION

With modern communication technology, air travel, and globalization, the world has never felt smaller. However, as barriers of space and time erode, a small-world view raises awareness of individual cultures, identities, and their interactions. This is referred to by UNESCO as a "resurgent affirmation of identities".[1] Culture directs how individuals see themselves and helps locate the groups with whom they identify. Language is important for cultural identity but creates barriers to communication. People who speak different languages require a bridge or link so one can understand the other. Solutions include learning each other's language (bilingualism) and communicating in a different shared foreign, or third, language. English is the most used second language in the world and is the most spoken language in the world [17] with twice as many speaking it as a second or third language than speak it as their first. In this way it has become a global language and there is great enthusiasm across many parts of the world to become proficient in its use. Whilst opening up communication across borders and lands is beneficial, one challenge with English is that its use internationally is seen by some as a form of cultural imperialism which erodes local identities and imposes a particular world view [8]. A second challenge with English as a foreign (third) language is that it creates an imbalance towards those with English as a first language and potentially dis-empowers non-native English speakers [6].

We are interested in communication using written (typed) text where the classic solution to different language communication is translation. Before machine translation (MT) was available, the disadvantages of translation were (a) the expense of good translators and (b) the time lag for translation [58]. But, MT has mitigated both these issues and is now widely used in business, leisure, and healthcare settings, where it enables communication between individuals with no shared language [28].

---

[1]Unesco.org

Machine translation has difficulty in accurately translating meanings – especially with unstructured or poorly formed text and with minority languages [2]. Critics of MT also note the bias towards English and a potential for misuse [59]. Even so, MT has revolutionized the way adults manage language. Common examples include the translation of menus [19], social media [11], and websites [24].

Machine translation is not widely studied with children, but studies from recognition systems find that children are sympathetic to errors and poor or unusual translations [45, 46]. This is helpful as studies in text entry show that children make many errors when they write, and errors do impact translation [26, 27]. A positive side to MT for children in a chat-style application is that it occurs in real time, thus improving engagement. In addition, studies with adults suggest that errors in translation create a space for invention and creativity and thereby keep a conversation going on account of the ambiguity [31, 48].

Our work seeks to facilitate communication between children from different cultures and with different languages. We do not want this communication to erode cultures nor to provide lop-sided interaction. The study reported here, bringing children from the UK and Nepal together, aims to explore the possibilities for machine-translated text in a chat application for children. We explore current problems, offer possibilities for improving interaction, and highlight the potential of MT to bring children across the world together.

## 2 BACKGROUND AND RELATED WORK

### 2.1 Children Writing

Children's first communication is oral. When they first use a pen, it is as a tool that makes marks with no association with spoken words. As children approach pre-school, they explore the equivalences of spoken and written language [21], and then, with improved skill in using a pen and other writing tools, they explore the symbolic properties of spoken and written language [16]. By the middle of childhood, their speaking and writing become differentiated: Written texts lengthen and start to conform to the "rules" of writing [14]. Writing is a complex task that requires motoric mastery, knowledge of the symbols, language construction, and adherence to grammatical rules [4]. Children can generally say much more than they can write as writing is fraught with challenges, especially in regard to spelling norms. It known that children change their intention in writing when, for example, they cannot spell a word [43]. Whether writing with a pen or a keyboard, spelling is a challenge for children [15].

Text entry (text input) is the process of entering text in a computer so it can be encoded (e.g., in Unicode) and then stored, manipulated, or presented [32]. Originally done with physical keyboards, today there are numerous methods including soft keyboards, as on smartphones, and recognition-based methods like speech and handwritten input. In all cases, text entry performance varies with experience [33]. Of the (numerous) text entry errors children make, some are a consequence of an inability to spell words [10], while others arise from mechanical errors like slips. It is important to note that without knowing a child's intention, it is difficult to know if an error is a slip at the keyboard or a spelling error [27].

### 2.2 Nepalese Text Entry

Nepalese is the spoken language of Nepal and for written texts it adopts the Devanagari script, one of several scripts derived from Brahmic script, which, as early as the 11th century, became the default script for writing Sanskrit [7]. Brahmic scripts have consonants that follow an alphabetic style but vowels positioned before, above, and around consonant collections in a syllabic style. Of the four main writing systems (scripts) across the world, the Latin script (alphabetic) is the only one for which the Qwerty keyboard is well suited. Non-alphabetic scripts like Logographic (e.g., Chinese), Abjad (e.g., Hebrew) and Abugida (e.g., Hindu) are poorly suited to standard keyboards but soft keyboards, with the potential for layering of menus and prediction, provide an option for them and have been explored within HCI [5, 34, 55]. Whilst there is specific research in HCI and elsewhere to build text input systems for Brahmic scripts [25, 34], with some using pen-based input [9], these initiatives are seldom commercialised. Consequently, most people using non-alphabetic scripts use phonetic writing systems.

Phonetic writing systems use symbols to enter the "sound" of a word; they are commonly used for non-Latin scripts (e.g., Brahmic). Whilst some languages have standardized phonetic equivalents (e.g., Chinese Mandarin Pinyin [35]), most phonetic text entry is not bound by rules in the same way that native scripts are. That is to say, "spelling" rules are not applied. In these instances phonetic writing is in essence a transliteration of spoken language – bringing with it all the variations that existed before written standards were imposed [3]. Such phonetic entry is fine so long as both writer and reader see the same thing. An example from English is a phonetic typist writing *"We wer out"* which could be interpreted as *"We wear out"* or *"We were out"*. Children are known to use phonetic text entry in their interactions with mobile phones: For them, the various ways of "sounding out" a word are natural and relatively unstudied, but it is expected that this might impact machine translation.

### 2.3 Machine Translation, Children and Writing

Machine translation (MT) of language was first suggested in the early 1930s [60]. With the advent of AI, translation software is now highly available and adaptive. Google Translate (GT), first introduced in 2006, is the most-widely used translation engine and currently translates between 100+ languages. The translation is sentence-level, so when words are translated that do not make sense in context, one or more of the words are altered in order to create a meaningful sentence; this makes it ideal for "faithful" as opposed to "literal" translation [39]. Faithful translation retains the contextual meaning of the original whereas a literal translation is word-for-word translation where the target language copies the source exactly; this can result in clunky text with poor flow. GT has been evaluated in a range of academic papers, but these primarily target adult-generated text. Examples include evaluations of the effectiveness of GT for English into Bahasa [37] and from English into Arabic [23].

Using machine translation with phonetic text input generally requires a pre-translation process whereby a typist enters phonetic text and then accepts a native language written version of the intended words – as suggested by a language-matching program. An example is Baraha which converts phonetic typing into Indic

script.[2] To type the word "kal", the user types KAL on a Qwerty keyboard and the word **कल** is displayed. This is referred to as transliteration – the process of expressing the sound of how a word is pronounced in the source language using the alphabet of the target language [1]. Research on Google Translate tends to focus on the errors, for example, "defective translation" [50] or gibberish [53].

Studies of Google Translate with children are rare, but tend to take a broader view. One study looked at GT's use in the classroom to facilitate learning [13]. The usefulness of translation, as applied to children and their language development and socialization, is the main theme of our own work. In the context of children's chatter, translation accuracy is not so important. The translation simply has to communicate essentials [38]. When nothing is especially essential, MT can be a tool for creativity and play. Literary experimentation is a playful approach where text is put into a translation tool with no expectation of a "correct" result [34].

In exploring cross-cultural chat with children, facilitated by machine translation (MT), we consider the ambiguity and uncertainty of automatic translation as something to work with rather than work against. Where there is uncertainty in a message, we are interested in how that might provoke a reply and keep the dialogue going. Our work seeks to instantiate machine translation in a pen pal application for real-time translated chat [12, 18].

## 3 THE STUDY

The aim of the present study is to explore possibilities for machine-translated text in a chat application for children. There were three research questions:

- RQ1 – How effective is GT at managing children's writing?

- RQ2 – What features of children's text impact translation?

- RQ3 – How do children react to MT text and does such hinder or help communication?

We answer RQ1 by studying accuracy and understandability of text translated from English into Nepalese. By correcting errors in inputted text we start to see how these errors impact on translation which contributes to RQ2. In a field study with children chatting online across two countries, we explore problematic words and symbols – looking at translations across both languages – this completes our work on RQ2. RQ3 addresses the context of example conversations that show how children appear robust and determined in the face of mistranslations. Ethical clearance for the study was obtained in the UK and Nepal.

### 3.1 Participants and Location

The work took place in the UK and Nepal with children aged 9 to 12. The first phase used writing gathered from 22 children from the UK with 36 children from Nepal reading that writing. The second phase, a field study, had 20 UK children and 38 Nepalese children engaging in online – almost synchronous – chat that was facilitated by Google Translate. Children's gender was not collected

---

[2]https://baraha.com/v10/index.php

but our samples included boys and girls. All the children were in mainstream school and were proficient in their own languages.

### 3.2 Apparatus

Children in Nepal used six identical Android mobile phones (Cubot Magic, with 5" display) and the children in England used six identical Android tablets (Samsung Galaxy Tab E, with 9.6" display). The bespoke Digipal app, built by the first author for this study, was preloaded onto the devices. Version 1 allowed children to enter text and then submit it to be translated and subsequently presented translated text to explore understandability. Version 2 live-translated text and showed it immediately to children in the second country; both versions included a built-in translation system (integrated Google Cloud translation API v3beta1) allowing each child to send and receive text in their own language. In the Nepalese version, the components (titles, labels, headers, etc.) on the app, as well as any received text, were shown in Nepalese. The Nepalese children used the Baraha keyboard which received phonetic input in a Latin script and translated it to Nepalese. Both the original and translated text were saved in a database along with sender and receiver IDs.

### 3.3 Procedure Phase 1 – Accuracy

Children in the UK worked in small groups in the University and constructed short letters to send to children that they had never met in Nepal. The letters the UK children wrote were converted into Nepalese by a Nepalese speaker to give what we refer to as a correct translation (CT). At the same time the letters were put through Google Translate to result in a Google translated version (GT). To explore the effect of children's spelling mistakes and input errors on accuracy, each English letter was then corrected for grammar and then translated again using Google Translate to give a third string, GTC. We used a variation of the MSD (minimum string distance) statistic [52] to compute accuracy. This formula finds the error rate of Google translated text with small modifications whereby entered text is replaced by Google Translated text (GT) and presented text is replaced with by correct (human translated) Text (CT). Hence, the equations are

$$Error\ Rate,\ ER = \frac{MSD(GT, CT)}{|CT|} \times 100\% \qquad (1)$$

and

$$Accuracy,\ A = (100 - ER)\% \qquad (2)$$

*MSD(GT, CT)* is the minimum string distance: the number of edits (insertions, deletions, substitutions) to convert *GT* into *CT*. |*CT*| is the number of characters in *CT*.

We used a small script to compute accuracy; however, on porting the data into the script, where a synonym was found, this was manually adjusted prior to the calculation. This is standard practice in evaluating machine translation, taking the same approach as Papineni et al. [41]. Examples that were found (and resolved) included dad vs. father, mummy vs. mother, my school's name vs. name of my school, and class vs. grade.

## 3.4 Results: Accuracy of Google Translation from English to Nepalese

The accuracy of the Google translation of the English letters was calculated using Eq. 2. Table 1 shows the full results for the 22 letters, where,

- |CTE| = number of characters in correct (human) translation

- |GTE| = number of characters in Google translation

- A = accuracy (see Eq. 2)

- % Gain = increase in accuracy from uncorrected to corrected

It is clear that the MSD edit distance varies by letter, with accuracy varying from 39.5% to 68.6%. The average is 54.8% but the high standard deviation of 6.97 indicates the data are scattered out from the mean which is probably a feature of children's writing. Even after correction for grammar, none of the translations were 100% accurate. Statistically, the results before cleaning ($M$ = 54.8, $SD$ = 7.0) and after cleaning ($M$ = 65.1, $SD$ = 8.5) indicate a significant improvement in accuracy, $t(21)$ = −7.36, $p$ < .00001.

## 3.5 Procedure Phase 1 – Understandability

Measuring understandability is complex. Readability is one metric that can be applied to text as a predictor of the information successfully conveyed to a larger population when people are trying to access it [40]. Readability is often seen as a proxy for understandability, but readability is independent of the human reader, it is a predictive measure with well understood metrics such as the Flesch reading ease metric [29].

Comprehension can serve as a proxy for understandability, where comprehension is measured in a variety of ways. An example is Taylor's Cloze (from "closure") technique, where a text is read and then a reader "fills in" missing words relying on contextual factors in the text [56]. Retelling a story is another method used to measure understanding [49]. Critiques of retell highlight that students have to recall information, organize it, then possibly draw conclusions; but, with short texts, retell does not suffer from these limitations [30, 47]. Retell is scored in many ways but the most common is to count idea units that were recalled and to present these as a proportion of idea units that were presented (e.g., [36, 54, 57]).

With translated text, there is no confirmed method to quantify understandability. Hassani [22], asked readers to rate the understandability of text using terms like "slightly understandable". Rossetti and O'Brien [51] used readability scores as well as retelling (as described above) to measure understandability of machine translated text with teenagers. Given the lack of a single standardized technique, we chose to combine a retell method with an idea-counting metric to determine understandability. The letters the children in the UK wrote each generated two translations, one from uncorrected English (GT) and one from corrected English (GTC). We considered understandabilty from two angles: predicted understandability and experienced understandability. In line with Miller and Keenan [36] and others, we accept that a sentence might contain one or more key pieces of information and that understandability could be measured by counting this information and seeing how much of it was understandable to the child after the translation process. Given that we were working with children's writing, and that the writing is often error prone, even at the point of entry, we determined a "predicted" understandability for each of the 22 letters based on the extent it made sense to an adult reader.

Experienced understandability was measured by asking children to inspect what they saw and to describe this to a third party drawing from Reed and Vaughn [47] and Roberts et al. [49]. For this part of the study, children in Nepal were allocated letters to read (half were from corrected text and half from uncorrected text) such that each UK letter (having been translated into Nepalese) was read by three or more Nepalese children. To explore the effect of children's spelling mistakes and input errors on understandability, comparisons were made across the scores for understandability from these corrected and uncorrected outputs.

The Digipal app was adapted to display the translated children's letters and prompt the children to describe the contents of these letters to a third party. The children looked at as many different letters as they could in 10 minutes, the letters were presented in the app in such a way that all the letters were studied roughly the same number of times. The children were asked not to copy the same text but to write what they understood instead. In determining if a child understood a piece of information, we happily accepted synonyms as being matched (as per [51]), and we ignored spelling mistakes and small grammatical mistakes.

For each text, we considered how much information it included and made a count. We refer to IO as the number of individual items of information in the original writing. In this example, "My name is Jack. I live in England, and I am 7 years old." there are three items so IO = 3. Once a letter was translated, the translation might have fewer items of meaningful information, as seen in this example:

Original text in Nepalese: मेरो नामे दिवस भुजेल हो मा क्लास ६ मा पद्रथु तमि मा अद शोउ

Correct translation: My name is Diwas Bhujel. I am in class 6. Which one are you in?

Google translation: My name is Bhujel ho in class 6 how much is it to show?

Here, IO = 3 as the original contains three pieces of meaningful information: name, school class, and a question. The Google Translation appears to have only two pieces of meaningful information as "how much is it to show?" makes no sense. This "predicted" number is referred to as IT (information in translated version).

## 3.6 Results: Predicted Understandability of Translated Letters

See Table 2. In the table,

- IO = number of items of meaningful information in original letter (before translation)

- IT = number of items of meaningful information in translated letter

- ITC = number of items of meaningful information in translated corrected letter

- % Gain = increase from uncorrected to corrected

Table 1: MSD edit distance, accuracy, and % gain of Google Translation of English letters

| | Human Translation | Machine Translation | | | | | | |
| | | Uncorrected | | | Corrected | | | |
| Letter | \|CTE\| | \|GTE\| | MSD | Accuracy | \|GTE\| | MSD | Accuracy | % Gain |
|---|---|---|---|---|---|---|---|---|
| E1 | 114 | 91 | 40 | 64.9 | 115 | 24 | 78.9 | 21.6 |
| E2 | 176 | 198 | 76 | 56.8 | 177 | 62 | 64.8 | 14.0 |
| E3 | 218 | 223 | 77 | 64.7 | 217 | 24 | 89.0 | 37.6 |
| E4 | 250 | 226 | 113 | 54.8 | 228 | 101 | 59.6 | 8.8 |
| E5 | 332 | 240 | 201 | 39.5 | 284 | 140 | 57.8 | 46.6 |
| E6 | 137 | 109 | 60 | 56.2 | 126 | 53 | 61.3 | 9.1 |
| E7 | 286 | 242 | 130 | 54.5 | 260 | 120 | 58.0 | 6.4 |
| E8 | 96 | 70 | 49 | 49.0 | 83 | 48 | 50.0 | 2.1 |
| E9 | 232 | 217 | 95 | 59.1 | 235 | 67 | 71.1 | 20.4 |
| E10 | 118 | 103 | 37 | 68.6 | 110 | 29 | 75.4 | 9.9 |
| E11 | 317 | 270 | 181 | 42.9 | 297 | 137 | 56.8 | 32.4 |
| E12 | 376 | 342 | 168 | 55.3 | 353 | 148 | 60.6 | 9.6 |
| E13 | 498 | 525 | 266 | 46.6 | 508 | 174 | 65.1 | 39.7 |
| E14 | 109 | 90 | 54 | 50.5 | 97 | 38 | 65.1 | 29.1 |
| E15 | 222 | 192 | 97 | 56.3 | 192 | 80 | 64.0 | 13.6 |
| E16 | 256 | 199 | 110 | 57.0 | 222 | 94 | 63.3 | 11.0 |
| E17 | 389 | 321 | 191 | 50.9 | 335 | 123 | 68.4 | 34.3 |
| E18 | 412 | 411 | 174 | 57.8 | 414 | 142 | 65.5 | 13.4 |
| E19 | 141 | 121 | 62 | 56.0 | 132 | 52 | 63.1 | 12.7 |
| E20 | 151 | 115 | 74 | 51.0 | 131 | 42 | 72.2 | 41.6 |
| E21 | 247 | 227 | 94 | 61.9 | 230 | 85 | 65.6 | 5.9 |
| E22 | 206 | 185 | 99 | 51.9 | 196 | 90 | 56.3 | 8.4 |
| Average | 240.1 | 214.4 | 111.3 | 54.8 | 224.6 | 85.1 | 65.1 | 19.5 |
| *SD* | 110.6 | 112.0 | 61.0 | 7.0 | 108.6 | 45.2 | 8.5 | 13.5 |

As expected, meaningful information was lost when letters were translated. Correcting before translation resulted in more meaningful information being retained. Note that, for example, E12 began with ten items of information; one was lost when translated and then regained when the letter was corrected. This letter was written as follows "My name is Anon. I live with my brother, mummy,daddy and me I have a hamster and 2 fish. My favourite food is stew and barbeque pork ribs. My school is St Anon c of e primary school .I'm in year 5. My favorite sport is dancing and drama.We make money by lots of different ways . Our school is surrounded by countryside and shops" The poor punctuation resulted in lost information when translated which was restored in a grammar-corrected version (capitalisation, full stops, etc.).

## 3.7 Results: Experienced Understandability of Translated Letters

Each child's retell was examined by the research team and scored based on the count of the items of meaningful information retained from the original (English) letter. A score was calculated for each letter by taking averages from all the children that looked at the 44 (22 uncorrected and 22 corrected) versions of the letters – note that no child looked at the corrected *and* uncorrected versions of the same letter.

See Table 3. In the table,

- IC = number of items of meaningful information retold by children from uncorrected letters (as an average)

- ICC = number of items of meaningful information retold by children from corrected letters (as an average)

Note here that the columns show what percentage these represent as percentage of the original, as in written in English (IO), were retained and percentage of the predicted expert view of the translation (IT or ITC) were retained.

Of interest here is that in some cases the predicted understandability was lower than the experienced understandability – that is to say the children, on average, understood more than the grown ups had – this may be because the children were more willing to guess at meanings. Cleaning improved translation and thus improved understandability – but not by much. It is worth noting that over a third of the letters retained all the meaningful information from start through to translation and to retell. This is encouraging.

## 3.8 Procedure Phase 2 – Conversation

For phase 2, two of the authors positioned themselves with children in two locations. After ethics clearance and the completion of consent processes with the children, parents, and schoolteachers in these venues, all the children were met one day before the activity to ensure they fully understood what they would be doing the next

**Table 2: Number of items of meaningful information conveyed (IO), remaining in the translated letter (IT) and in the corrected translated letter (ITC)**

| Letter | English IO | Uncorrected GT | | Corrected GT | | % Gain |
|---|---|---|---|---|---|---|
| | | IT | IT as % of IO | ITC | ITC as % of IO | |
| E1 | 4 | 3 | 75.0 | 4 | 100.0 | 33.3 |
| E2 | 7 | 6 | 85.7 | 6 | 85.7 | 0.0 |
| E3 | 8 | 8 | 100.0 | 8 | 100.0 | 0.0 |
| E4 | 7 | 7 | 100.0 | 7 | 100.0 | 0.0 |
| E5 | 10 | 9 | 90.0 | 9 | 90.0 | 0.0 |
| E6 | 3 | 3 | 100.0 | 3 | 100.0 | 0.0 |
| E7 | 8 | 8 | 100.0 | 8 | 100.0 | 0.0 |
| E8 | 3 | 3 | 100.0 | 3 | 100.0 | 0.0 |
| E9 | 7 | 7 | 100.0 | 7 | 100.0 | 0.0 |
| E10 | 4 | 4 | 100.0 | 4 | 100.0 | 0.0 |
| E11 | 9 | 8 | 88.9 | 8 | 88.9 | 0.0 |
| E12 | 10 | 9 | 90.0 | 10 | 100.0 | 11.1 |
| E13 | 15 | 15 | 100.0 | 15 | 100.0 | 0.0 |
| E14 | 5 | 5 | 100.0 | 5 | 100.0 | 0.0 |
| E15 | 6 | 6 | 100.0 | 6 | 100.0 | 0.0 |
| E16 | 7 | 7 | 100.0 | 7 | 100.0 | 0.0 |
| E17 | 11 | 9 | 81.8 | 10 | 90.9 | 11.1 |
| E18 | 12 | 12 | 100.0 | 12 | 100.0 | 0.0 |
| E19 | 5 | 4 | 80.0 | 5 | 100.0 | 25.0 |
| E20 | 7 | 7 | 100.0 | 7 | 100.0 | 0.0 |
| E21 | 8 | 8 | 100.0 | 8 | 100.0 | 0.0 |
| E22 | 5 | 5 | 100.0 | 5 | 100.0 | 0.0 |
| Average | 7.32 | 6.95 | 95.06 | 7.14 | 97.98 | 3.66 |
| *SD* | 3.01 | 2.94 | 8.01 | 2.92 | 4.47 | 8.97 |

day (which would be writing to a child in Nepal or England) and to ensure there was no uncertainty.

On the first day of the main activity, one researcher (the first author) and two volunteers began the study in the Banepa School around 09.00 local time (+4:45 GMT). Due to there being only six mobile devices available, the children took turns writing their letters. Each child was given a unique user ID (user 2, user 4, user 6, and so on) and asked to remember this for later. The children logged on and started writing their first letter, introducing themselves to the children in the UK. They were encouraged to ask one or more questions to the UK child in order to encourage a reply. Once a child finished, the writing was quickly checked (to ensure it was appropriate and enough words had been written), the child pressed send and he/she then passed the device to another child so that all the children were able to send their first letter.

Meanwhile in England, one researcher (the second author) and a volunteer teacher started the day at 09.00 local time (GMT). Due to the time difference, the letters from Nepal had arrived and so in England the children began by answering those letters. Six children at a time could see and reply to letters due to once again only having a limited number of mobile devices. These children had odd-numbered logins (user 1, user 3, user 5, etc.) such that user 1 and user 2 would write to each other, and user 3 to user 4, and so on. The children in the UK receiving letters were asked to read carefully and try to understand what was being said (even if there were mistakes).

Then they introduced themselves, answer any questions and ask what they wanted to know about the child who had written to them. Similar to the protocol in Nepal, once finished, the writing was briefly checked, and the child then sent the letter and handed the tablet to the researcher for another child to use.

At the end of day one in Nepal, all the children had written a first letter and half of them also got a chance to read the reply to it and write a second letter. By the end of day one in England, all the children had received one letter and written one letter as a reply. On day two, similar activities continued. At the end of this day all the children had exchanged at least two letters each. Some had exchanged three letters and one pair shared four letters each.

Whilst the earlier phases of this study had shown us that grammatical errors, spelling mistakes and entry slips all affected translation accuracy, the earlier work did not look individually, across naturally occurring chat-based text, to see where errors were made. Just as for text input, when we do not know the intent of a child, we cannot be sure whether they have made a mistake at the point of entry. Note that in this study, the Nepalese children wrote using a phonetic keyboard add-in to the app, from the Google Play Store and so spelling mistakes were unusual as they had to type the sound of a word and then the keyboard suggested words that "should" have been spelt right, As the child wrote a word in Latin script (as it sounded), the native Nepalese word appeared. If the word was wrong, other words were available to pick from. An error

**Table 3: Number of items of meaningful information retold by children for uncorrected (IC) and corrected (ICC) texts.)**

| | Uncorrected Understood | | | Corrected Understood | | | |
|---|---|---|---|---|---|---|---|
| Letter | IC | IC as % of IO | IC as % of IT | ICC | ICC as % of IO | ICC as % of IT | % Gain |
| E1 | 3.00 | 75.0 | 100.0 | 4.00 | 100.0 | 100.0 | 33.3 |
| E2 | 5.67 | 81.0 | 94.4 | 6.00 | 85.7 | 100.0 | 5.8 |
| E3 | 7.75 | 96.9 | 96.9 | 7.75 | 96.9 | 96.9 | 0.0 |
| E4 | 7.00 | 100.0 | 100.0 | 7.00 | 100.0 | 100.0 | 0.0 |
| E5 | 8.50 | 85.0 | 94.4 | 8.75 | 87.5 | 97.2 | 2.9 |
| E6 | 3.00 | 100.0 | 100.0 | 3.00 | 100.0 | 100.0 | 0.0 |
| E7 | 7.50 | 93.8 | 93.8 | 7.75 | 96.9 | 96.9 | 3.3 |
| E8 | 3.00 | 100.0 | 100.0 | 3.00 | 100.0 | 100.0 | 0.0 |
| E9 | 5.67 | 81.0 | 81.0 | 6.50 | 92.9 | 92.9 | 14.6 |
| E10 | 3.75 | 93.8 | 93.8 | 3.75 | 93.8 | 93.8 | 0.0 |
| E11 | 7.75 | 86.1 | 96.9 | 7.75 | 86.1 | 96.9 | 0.0 |
| E12 | 9.00 | 90.0 | 100.0 | 10.00 | 100.0 | 96.7 | 11.1 |
| E13 | 14.67 | 97.8 | 97.8 | 14.67 | 97.8 | 100.0 | 0.0 |
| E14 | 5.00 | 100.0 | 100.0 | 5.00 | 100.0 | 100.0 | 0.0 |
| E15 | 6.00 | 100.0 | 100.0 | 6.00 | 100.0 | 100.0 | 0.0 |
| E16 | 7.00 | 100.0 | 100.0 | 7.00 | 100.0 | 100.0 | 0.0 |
| E17 | 8.00 | 72.7 | 88.9 | 9.25 | 84.1 | 92.5 | 15.6 |
| E18 | 9.25 | 77.1 | 77.1 | 11.00 | 91.7 | 91.7 | 18.9 |
| E19 | 3.67 | 73.3 | 91.7 | 4.75 | 95.0 | 95.0 | 29.4 |
| E20 | 5.50 | 78.6 | 78.6 | 6.00 | 85.7 | 85.7 | 9.1 |
| E21 | 8.00 | 100.0 | 100.0 | 8.00 | 100.0 | 100.0 | 0.0 |
| E22 | 5.00 | 100.0 | 100.0 | 5.00 | 100.0 | 100.0 | 0.0 |
| Average | 6.53 | 90.09 | 94.78 | 6.91 | 95.18 | 97.09 | 6.56 |
| *SD* | 2.70 | 10.32 | 7.25 | 2.78 | 5.81 | 3.82 | 10.02 |

can occur if the phonetically typed word returned the wrong word in native Nepalese but the child does not realize it. For example, if the child meant to write "win" – for which the phonetic typing is probably "jit", but by writing "git" the Nepalese word for "song" appeared. If the child realized the error, he/she could selected the right word from the suggestion box. Another problem occurs if the phonetically typed word is wrong and does not result in a clean match in Nepalese. Then, the app might not return a Nepalese word at all. Because of this extra layer of input, it has to be noted that for Nepalese, there are three strings for each letter segment:

- IPT: inputted phonetic text
- CNT: converted Nepalese text
- GTE: Google translation to English text

Because we did not capture the input string at the app, we can only consider translation effectiveness at the level of CNT to GTE.

The translation from English to Nepalese was direct without any phonetic rendering introduced. Therefore, there are just two strings for each letter segment:

- IET: inputted English text
- GTN: Google translation to Nepalese text

In this phase we lined up inputted English text and converted Nepalese text with their translated equivalents and, iterating back and forth and using a codebook that evolved during the analysis, sought to understand WHAT had made the error rather than to

compute an error score. We were particularly interested to find which errors were common. We then looked at individual conversations across the two countries to explore the effect of errors on the conversation. We present the results in three main sections. First, we describe the composition and presentation of the texts and highlight our findings in relation to inputted children's text. Then, we look at the effectiveness of translation as afforded by Google before finally exploring the effects on conversation and flow.

### 3.9 Results: What was written

There were 190 letters exchanged, consisting of 111 Nepalese letters and 79 English letters. All the children exchanged at least two letters each. Some exchanged three letters, and one pair managed to send four letters each. Spelling mistakes were seen in about 20% of the letters – examples included "defrent" for "different", "lot's" for "lots" and "over" for "other". Sometimes children missed putting in spaces which resulted in non-words, for example, "Xboxand" and "pcand". We accept that in assuming the text is not what the child intended, we are making some judgments [32]. We assume here that the child intended to write "but" not "buy" (most probably a typing slip given the proximity of *t* to *y* on the keyboard):

- IET: I used to have a dog buy he died.

- Google output: म एउटा कुकुर किन्नको लागि प्रयोग गर्थे ऊ मरेको थियो

- HTE: I used to use a dog for buying he had died

In this second example, a probable spelling error simply resulted in Google deleting that word.

- IET: What **relogian** do you have?

- Google output: तपाईसँग के छ?

- HTE: What do you have?

In both these cases, which were typical, it is clear that the writing seen by the recipient did not convey the meaning as intended. Input slips creating unintended words and spellings were the ones most likely to affect the success of Google Translate. Additional words that are small and missing punctuation had little effect. Given that Google does smart translation a small mistake in the child's writing could have a significant effect on the meaning conveyed.

## 3.10   Results: Translation

*3.10.1   English to Nepalese.* Here, due to space limitations, we highlight only those errors encountered on multiple occasions. Number translation was problematic and Google Translate often (8 instances) deleted the number as illustrated.

- IET: Written text: I am **9 years old**.

- GTN: Google output: म **years बर्षको** भएँ। (HTE: I am **years old**.)

Pronouns were not always correctly translated, as in the following example:

- IET: Written text: What food do **you** like

- GTN: Google output: **मलाई** कस्तो खाना मनपर्छ (HTE: What food do I like)

Place names and proper words were problematic (7 instances). The word "England" stayed in Latin script in the Google translation as shown here:

- IET: Written text: I live in England.

- Google output: म इ England ्ल्याण्डमा बस्छु। (HTE: I live in Englandland)

When a "proper" name also had a non-proper meaning – the name was translated not as a name and therefore did not make sense in the context.

- IET: Written text: My teacher's name is miss pickles.

- GTN: Google output: मेरो शिक्षकको नाम मिस मिसिएको अचार हो (HTE: My teacher's name is miss mixed pickles)

*3.10.2   Nepalese to English.* Errors seen were similar to those in the English letters. There were 92 instances when Nepalese numbers (written in digits) were translated to a different number than intended.

- CNT: Written text: म १३ वर्ष को भएँ।

- HTE: Expected output: I am **13** years old.

- GTE: Google output: I am **3** years old.

There were also problems with personal pronouns probably as there are no gendered pronouns in Nepalese and GT was not able to figure which one to match based on the context.

- CNT: Written text: मेरो **आमाको** नाम जानुका तिमल्सिना हो। **उहाँ** ३२ वर्षको हुनुभयो।

- HTE: Expected output: My **mother's** name is Januka Timalsina. **She** is 32 years old.

- GTE: Google output: My **mother's** name is Januka Timalsina. **He** is 3 years old.

Nepalese months were translated wrong every time (eight in total), possibly due to the different calendar system in Nepal and to Google Translate's inability to map them correctly.

- CNT: Written text: मेरो जन्म **भदौ** २९ गते भएको हो ।

- HTE: Expected text: I was born in **Bhadau** 29.

- GTE: Google output: I was born in **August** 29. (NOTE: which is wrong)

Many Nepalese people or place names were translated incorrectly (11 instances) with the output giving a meaning associated with the name, rather than the name itself.

- CNT: Written text: मेरो नाम **विशाल पाठक** हो ।

- HTE: Expected output: My name is **Bishal Pathak**.

- GTE: Google output: My name is **huge reader**. (NOTE: both name and surname have meaning)

An interesting case of a Nepalese word being incorrectly mapped wrong was "momo", this is a popular food that was frequently mentioned in Nepalese children's letters. There were 22 times where this noun was translated to a variety of terms in English, for example, "me", "mum", and none of them fit properly in the context. This change of meaning resulted in very strange sentences, such as "my favourite food is mum", but the receivers continued the conversations by echoing something similar. For example:

- CNT: Written text: मेरो मनपर्ने खाना **मःमः** हो।

- HTE: Expected output: My favourite food is **momo**.

- GTE: Google output: My favourite food is **mum**.

- Reaction: My favourite food is tuna.

We postulate that one reason "momo" was translated in so many different ways was because it was unknown to the translation software and so the software made a guess. This is both a drawback and in other cases a strength of a smart translation application. Sadly, the software never managed to substitute a food name for "momo".

## 3.11   Results: Conversations

The letters in both languages, though "written" were actually short (on average 51.1 words) and casual. We can consider them "written conversations" or even "transcripts" of children's conversations, though the "transcription" was done by children themselves. Letter conversations usually started with identity of speakers, such as names. Children were aware of the social functions of taking turns,

and often explicitly used prompt questions to invite the hearer (recipient) into the conversations. Looking at this example we can see how the conversation continued even with breakdowns of meaning. We note – see emboldened text – that the question from letter two about the national animal is answered in letter three, and that the stream about a dog is continued through the letters. Teams (assumed to be football) are also carried through the conversation. Note, to protect identities, we substitute for original names.

(1) IET: Hi my name is Danny I live in England and have a big house it is very cold here but very nice I live in a big town there is a shop down the road to buy stuff and a big park it is nice talking to you bye bye

(2) MTE: I am very happy that the **Dhanbad team** has said this. **I have a yacht pet dog**. Ma Chai is from Nepal. The name of the mine is my favorite thing is ma'am. The name of **our national animal** is the cow. What is the national animal of your country?

(3) IET: the **national animal is a lion** I used to **have a dog** buy he died so I'm going to get a Shiva puppy my favourite **football team is Liverpool and I hate Manchester** they are rubbish my favourite things are my Xbox my phone my toy gun my Nintendo switch my teddy and blanket and my family I have a bestist best friend called John Smith nice talking to you see you later bye bye.

(4) MTE: I do **not have a dog** slaughtered even once. Name of your Facebook ID. I am capable Thapa. I like your face. My face is fine. **How is your dog**? I'm fine.

Translation sometimes created confusion and resulted in a question which would open up a conversation. The confusion is resolved and the two are amicable.

(1) My favourite pet is a cat. (MTE: My favourite belly is a cat)

(2) Do you eat cat? Do you want to be best friend?

(3) No I like cat. Yes I want to be friend.

Humour is seen in several of the exchanges and comes directly from the translation:

(1) Do you watch logan paul? (MTE: are you logan paul?)

(2) I am not logan paul. (MTN: I am logan paul)

(3) Why are you Joking, you are not logan paul.

From this example it is seen that the children were robust enough to deal with translation errors which provoked questions, lead to humour, and created divergence. The key observation from looking at all the children's conversations was that translation errors did not interrupt the process and often supported continuation.

# 4 DISCUSSION

## 4.1 Methods to study Translation with Children's Text

In our work we have applied two measures, one for accuracy and one for understandability. Our position within this paper is that we consider the measures to be appropriate but do not claim them to be optimal. In choosing a CER measure for accuracy we accept that this has limitations. Accuracy measures for translation are relatively new but in line with the assumption that translation can never be entirely accurate, many measures currently used apply a variety of layers to ensure that when a translation moves

words around or loses some local idioms, it can be automatically scored. Understandability was measured using retell, mainly as this was easy for children to do and had, as per [47] and others, been reported as suitable for use with children. In future studies we would suggest a triangulation of methods for understandability as it could be argued that in our case our measure was relatively coarse grained. Given the short length of the letters in our case, retell was manageable. Our intention in measuring understanding was also to apply a method that could be used in a classroom that limited the possibilities to confer and chat about what the children had read. At the same time we wanted a relatively fun activity and it was observed that the children reading the letters, who were in Nepal, enjoyed using the technology and were keen to read as many letters as possible. We have highlighted the tension around whether understandability should be computed aligned to the child's initial writing or to the 'predicted understandability' of the translated text. We argue that the latter of these is a more appropriate measure of the effectiveness of the translation software than the former, therefore, with children's text, we recommend that predicted understanding should be computed.

## 4.2 Answering the Research Questions

*4.2.1 How effective is GT at managing children's writing?* Whilst 'accuracy rates' were low ($M = 54.83$, $SD = 6.97$), these were improved by cleaning the inputted text in the way a simple grammar / spell checker might clean ($M = 65.1$, $SD = 8.5$). Understandability was better preserved in the main with rates of ($M = 90.1$, $SD = 10.32$) when compared with uncorrected children's inputted text rising to ($M = 97.1$, $SD = 3.82$) when compared with grammar corrected text with nonsensical information uncounted. These figures point to the gains that can be made when children's writing is corrected before translation whilst also recognising that some of children's writing will be without meaning. Whilst not unique to children, the field study showed how Google Translate had difficulties with proper nouns, pronouns, and numbers. The faulty translation of months was unexpected and probably noticed due to the significance of birthdays and ages to children. Similarly, the confusion over "momo" was child-related and is unlikely to occur in business communication. It was something GT had just not learned. For those designing MT, we recommend more work on collecting corpora of children's writing and potentially having options for age-targeted translations.

*4.2.2 What features of children's text may impact translation?* Children made many mistakes. Input slips creating unintended words and spellings clearly affected the success of Google Translate. The field study, and the understandability scores, highlighted that missing punctuation had little effect on overall meaning and also showed how small words could be lost or gained without too much concern as children could read between the lines. This is referred to in the literature as "gisting" [20], and is common with translation systems – the child only needs to make sense of what they see. In a few cases, given Google's smart translation, a single mistake could have a significant effect on meaning. Auto spellcheck could help avoid some, but not all, of these problems. Children could be alerted to wrongly typed words by adding a speak-out option before sending messages. Some of the words used – names of local

food "momo", and many of the types of words used; place names, family names, months, and ages, appeared to exacerbate translation problems. This suggests that the content of children's writing, as well as the product, has to be considered. Until such time as MT includes options for child content, it is wise to build systems that encourage spelling out words and parenthesize proper nouns, so they stay untranslated.

*4.2.3 How do children react to MT text? – Do such texts hinder or help communication?* In answer to RQ3 we note that machine translation did not hinder communication. Where translations were strange, this did not break the communication; indeed, it often kept the conversation going. We see this as a good example of prismatic translation [44], where translation opens rather than closes and in which errors become a creative experience. Children's translated texts must be checked, since it is problematic for children to see auto-translated text without an adult filter. This is possibly something that needs building into any interface.

## 4.3 Children's Experiences

We cannot conclude the discussion on this paper without considering the experience of the children and the potential that MT brings to cross-cultural inter-lingual conversation for children. The initial study held the interest of the children in both locations – writers in the UK, readers in Nepal, but the final study was entirely different. In the two venues the children were visibly and audibly excited when they found that they had had a "response". Reading that first response, and replying to it, was probably the first time any of these children had communicated directly, in their own words, with someone so far away. There was chatter and laughter as they tried to make sense of what had been sent to them. It was a joy to hear the children share, in spoken words, what their own "twins" abroad had sent hem. Whilst there are e-systems that allow children to talk with one another in different languages, these are primarily aimed at developing second language skills rather than child-child communication and the preservation of local language [42].

## 5 CONCLUSIONS AND FURTHER WORK

We described a study across two countries with children communicating in their own languages and facilitated by Google Translate. Translation accuracy was measured before and after grammar and spelling correction of children's writing, and understandability of translated text was measured using retell by children receiving the translated text and by comparing this with the original text. In a field study the causes of low accuracy in translation and the effect of misunderstandings were explored and this showed that children could continue conversations in spite of, or even along with, misstranslation. Some errors were due to differences in the structure of the languages (e.g., pronouns) and others by the absence of proper names in the dictionaries. This confirms criticisms that translation software favours Western alphabets and Western languages [60]. The multiple different translations of "momo" exposed GT's use of context when word-for-word fails. This had the unintended consequence of implying one of the children had reverted to cannibalism! Translation software, when applied with children, must be handled with care, as with children searching the Internet [15].

We developed methods to use with children to explore translation software and reported quantitative and qualitative data that are useful as a benchmark in future studies. The choice to study English to Nepalese language was borne out of the first authors' desire to see Nepalese remain the language of the Nepalese people in a world where English pervades. The relative success of the field study with children communicating synchronously with Google Translate as a facilitator is encouraging for further similar systems. Nepalese is not one of the most used Google languages and this probably helped expose limitations in automatic translation for less common languages used around the world. In addition, this study used children's writing which is itself problematic for systems built with adult users in mind.

Having shown that error correction at the outset of text input can improve both accuracy and understandability we would recommend this be incorporated into children's translation interfaces. We would also recommend that the input interface include a way of highlighting proper nouns that should not be translated. For machine translation experts, we encourage a possible children's setting that might favour their "sort of" language when applying smart translation. Our own further work is expected to include studies to measure the effectiveness of these approaches alongside design studies to explore ways to keep engagement with the interface high in order to sustain interaction.

## 6 SELECTION AND PARTICIPATION OF CHILDREN

Schools were contacted and the schools who agreed to participate then chose which children would participate. There was no filtering of children; we worked with the entire group that had consented and were present. In both venues all children had the language of the locality as their first language. Ethics clearance was obtained for both locations with each location staffed by one of the authors and the class teacher. The children had the study clearly explained to them and so, as well as having consent from adults, they also agreed to take part.

## REFERENCES

[1] Md Faruquzzaman Akan. 2018. Transliteration and translation from Bangla into English: A problem solving approach. *British Journal of English Linguistics* 6, 6 (2018), 1–21.
[2] Aria Septi Anggaira. 2017. Linguistic errors on narrative text translation using Google Translate. *Pedagogy: Journal of English Language Teaching* 5, 1 (2017), 1–14.
[3] Mark Aronoff. 1994. Spelling as culture. In *Writing systems and cognition*, C. W. Watt (Ed.). Springer, Berlin, 67–86.
[4] Carl Bereiter and Marlene Scardamalia. 1983. Does learning to write have to be so difficult? In *Learning to Write: First Language / Second Language*, Ian Pringle, Janice Yalden, and Avia Freedman (Eds.). Rutledge, New York, 20–33.
[5] Xiaojun Bi, Shumin Zhai, and Barton A. Smith. 2012. Multilingual touchscreen keyboard design and optimization. *Human-Computer Interaction* 27, 4 (2012), 352–382.
[6] James J. Bradac. 1992. Thoughts about floors not eaten, lungs ripped, and breathless dogs: Issues in language and dominance. *Annals of the International Communication Association* 15, 1 (1992), 457–468.
[7] William Bright. 1996. The Devanagari script. In *The world's writing systems*, Robert T. Daniels and William Bright (Eds.). Oxford University Press, New York, 384–390.
[8] Martin Carnoy. 1974. *Education as cultural imperialism.* Longman, New York.
[9] Yuanyuan Chen, Weilan Wang, Tiejun Wang, and Caike Zhaxi. 2019. Multi-function Tibetan input method on Android. In *International Conference on Systems and Informatics – ICSAI 2019.* IEEE, New York, 643–648.

[10] William E. Cooper. 1983. *Cognitive aspects of skilled typewriting.* Springer, New York.

[11] Lorenzo Coretti and Daniele Pica. 2018. Facebook's communication protocols, algorithmic filters, and protest. In *Social media materialities and protest: Critical reflections*, Thomas Poell Mette Mortensen, Christina Neumayer (Ed.). Rutledge, London.

[12] Elizabeth O. Crawford and Misty M. Kirby. 2008. Fostering students' global awareness: Technology applications in social studies teaching and learning. *Journal of Curriculum and Instruction* 2, 1 (2008), 56–73.

[13] Diane Dagenais, Kelleen Toohey, Alexa Bennett Fox, and Angelpreet Singh. 2017. Multilingual and multimodal composition at school: ScribJab in action. *Language and Education* 31, 3 (2017), 263–282.

[14] Colette Daiute. 1989. Play as thought: Thinking strategies of young writers. *Harvard Educational Review* 59, 1 (1989), 1–24.

[15] Allison Druin, Elizabeth Foss, Leshell Hatley, Evan Golub, Mona Leigh Guha, Jerry Fails, and Hilary Hutchinson. 2009. How children search the Internet with keyword interfaces. In *Proceedings of the 8th International Conference on Interaction Design and Children – IDC 2009.* ACM, New York, 89–96.

[16] Anne Haas Dyson. 1991. Viewpoints: The word and the world: Reconceptualizing written language development or do rainbows mean a lot to little girls? *Research in the Teaching of English* 25, 1 (1991), 97–123.

[17] Ethnologue. 2023. Languages of the world. https://www.ethnologue.com/ [Online; accessed Mar 22, '23].

[18] Simeon David Flowers. 2015. Developing intercultural communication in an ELF program through digital pen pal exchange. *The Center for ELF Journal* 1, 1 (2015), 25–39.

[19] Adrián Fuentes-Luque. 2017. An approach to analysing the quality of menu translations in southern Spain restaurants. *Journal of Multilingual and Multicultural Development* 38, 2 (2017), 177–188.

[20] Spence Green, Jeffrey Heer, and Christopher D. Manning. 2013. The efficacy of human post-editing for language translation. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems – CHI 2013.* ACM, New York, 439–448.

[21] Jerome C. Harste. 1984. *Language stories & literacy lessons (ERIC No. ED257113).* Heinmann Educational Books, Portsmouth, NH.

[22] Hossein Hassani. 2017. Kurdish interdialect machine translation. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects.* Association for Computational Linguistics, Philadelphia, 63–72. https://doi.org/10.18653/v1/W17-1208

[23] Omar Jabak. 2019. Assessment of Arabic-English translation produced by Google Translate. *International Journal of Linguistics, Literature and Translation (IJLLT)* 0 (2019), 2617–0299.

[24] Miguel A. Jiménez-Crespo. 2013. *Translation and web localization.* Routledge, London.

[25] Manjiri Joshi, Anirudha N. Joshi, Nagraj Emmadi, and Nirav Malsattar. 2014. Swarachakra keyboard for Indic scripts (tutorial). In *Proceedings of the 1st International Conference on Mobile Software Engineering and Systems.* IEEE, New York, 5–6.

[26] Akiyo Kano and Janet C. Read. 2009. Causes of simultaneous keystrokes in children and adults. In *Proceedings of the IFIP Conference on Human-Computer Interaction – INTERACT 2009.* Springer, Berlin, 137–140. https://doi.org/10.1007/978-3-642-03655-2_16

[27] Akiyo Kano and Janet C. Read. 2009. Text input error categorisation: solving character level insertion ambiguities using Zero Time analysis. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology.* British Computer Society, London, 293–302.

[28] Ravish Kapoor, Angela T. Truong, Catherine N. Vu, and Dam-Thuy Truong. 2020. Successful verbal communication using Google Translate to facilitate awake intubation of a patient with a language barrier: A case report. *A&A Practice* 14, 4 (2020), 106–108.

[29] J. Peter Kincaid, Robert P. Fishburne, R. L. Rogers, and Brad S. Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (no. RBR-8-75).* Technical Report. Naval Technical Training Command Millington TN Research Branch, Millington, Tenn.

[30] Janette K. Klingner. 2004. Assessing reading comprehension. *Assessment for Effective Intervention* 29, 4 (2004), 59–70. https://doi.org/10.1177/073724770402900408

[31] Katrin Kohl and Wen-chin Ouyang. 2020. Introducing creative multilingualism. In *Creative Multilingualism: A Manifesto*, Katrin Kohl, Rajinder Dudrah, Andrew Gosler, Suzanne Graham, Martin Maiden, Wen chin Ouyang, and Matthew Reynolds (Eds.). Open Book Publishers, Cambridge, UK, 1–24.

[32] I. Scott MacKenzie and R. William Soukoreff. 2002. Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction* 17, 2 (2002), 147–198.

[33] I. Scott MacKenzie and Shawn X. Zhang. 2001. An empirical investigation of the novice experience with soft keyboards. *Behaviour and Information Technology* 20 (2001), 411–418.

[34] Aleksandra Małecka and Piotr Marecki. 2018. Literary experiments with automatic translation: A case study of a creative experiment involving King Ubu and Google Translate. In *On the Fringes of Literature and Digital Media Culture*, Irena Barbara Kalla, Patrycja Poniatowska, and Dorota Michułka (Eds.). Brill Rodopi, Leiden, The Netherlands, 77–88.

[35] Catherine McBride-Chang, Dan Lin, Phil D. Liu, Dorit Aram, Iris Levin, Jeung-Ryeul Cho, Hua Shu, and Yuping Zhang. 2012. The ABC's of Chinese: Maternal mediation of Pinyin for Chinese children's early literacy skills. *Reading and Writing* 25, 1 (2012), 283–300.

[36] Amanda C. Miller and Janice M. Keenan. 2009. How word decoding skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia* 59, 2 (2009), 99–113.

[37] Melita Nadhianti. 2016. *An analysis of accuracy level of Google Translate in English-Bahasa Indonesia and Bahasa Indonesia-English translations.* Journal Article. Yogyakarta: Universitas Negeri Yogyakarta.

[38] Sarah A. Nagro. 2015. PROSE checklist: Strategies for improving school-to-home written communication. *Teaching Exceptional Children* 47, 5 (2015), 256–263.

[39] Peter Newmark. 1988. *A textbook of translation.* Prentice Hall, New York.

[40] Pawan Kumar Ojha, Abid Ismail, and Kuppusamy Kundumani Srinivasan. 2021. Perusal of readability with focus on web content understandability. *Journal of King Saud University - Computer and Information Sciences* 33, 1 (2021), 1–10. https://doi.org/10.1016/j.jksuci.2018.03.007

[41] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, Philadelphia, 311–318. https://doi.org/10.3115/1073083.1073135

[42] Beryl Plimmer, Liang He, Tariq Zaman, Kasun Karunanayaka, Alvin W Yeo, Garen Jengan, Rachel Blagojevic, and Ellen Yi-Luen Do. 2015. New interaction tools for preserving an old language. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems – CHI 2015.* ACM, New York, 3493–3502.

[43] Janet C. Read. 2002. Optimising text input for young children using computers for creative writing tasks. In *Proceedings of 16th British HCI Conference*, Vol. 2. British HCI Group, London, 3 pages.

[44] Janet C. Read, Stuart MacFarlane, and Chris Casey. 2001. Measuring the usability of text input methods for children. In *People and Computers XV—Interaction without Frontiers.* Springer, London, 559–572.

[45] Janet C. Read, Stuart MacFarlane, and Chris Casey. 2002. Oops! Silly me! Errors in a handwriting recognition-based text entry interface for children. In *Proceedings of the Second Nordic Conference on Human-Computer Interaction – NordiCHI 2002.* ACM, New York, 35–40.

[46] Janet C. Read, Stuart MacFarlane, and Chris Casey. 2003. Good enough for what?: Acceptance of handwriting recognition errors by child users. In *Proceedings of the 2003 Conference on Interaction Design and Children – IDC 2003.* ACM, New York, 155–155.

[47] Deborah K. Reed and Sharon Vaughn. 2012. Retell as an indicator of reading comprehension. *Scientific Studies of Reading* 16, 3 (2012), 187–217. https://doi.org/10.1080/10888438.2010.538780

[48] Matthew Reynolds, Sowon S. Park, and Kate Clanchy. 2020. Prismatic translation. In *Creative Multilingualism: A Manifesto*, Katrin Kohl, Rajinder Dudrah, Andrew Gosler, Suzanne Graham, Martin Maiden, Wen chin Ouyang, and Matthew Reynolds (Eds.). Open Book Publishers, Cambridge, UK, 131–150.

[49] Greg Roberts, Roland Good, and Stephanie Corcoran. 2005. Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly* 20, 3 (2005), 304.

[50] Mónica Rodríguez-Castro, Spencer Salas, and Tracey Benson. 2018. To Google Translate™ or not? Newcomer Latino communities in the middle. *Middle School Journal* 49, 2 (2018), 3–9.

[51] Alessandra Rossetti. 2019. Evaluating the impact of semi-automation on the readability and comprehensibility of health content. In *Translation in Cascading Crises*, Federico M. Federici and Sharon O'Brien (Eds.). Routledge, London, 22 pages.

[52] R. William Soukoreff and I. Scott MacKenzie. 2001. Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic. In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems - CHI 2001.* ACM, New York, 319–320. https://doi.org/10.1145/634067.634256

[53] Paul Stapleton and Becky Leung Ka Kin. 2019. Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong. *English for Specific Purposes* 56 (2019), 18–34.

[54] Zinar Susan. 1990. Fifth-graders' recall of propositional content and causal relationships from expository prose. *Journal of Reading Behavior* 22, 2 (1990), 181–199.

[55] Kumiko Tanaka-Ishii and Renu Gupta. 2010. Writing system variation and text entry. In *Text Entry Systems: Mobility, Accessibility, Universality*, I. Scott MacKenzie and Kumiko Tanaka-Ishii (Eds.). Morgan Kaufmann, San Francisco, 191–202.

[56] Wilson L. Taylor. 1953. "Cloze Procedure": A new tool for measuring readability. *Journalism Quarterly* 30, 4 (1953), 415–433. https://doi.org/10.1177/

107769905303000401

[57] Paul van den Broek, Yuhtsuen Tzeng, Kirsten Risden, Tom Trabasso, and Patricia Basche. 2001. Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology* 93 (09 2001), 521–529. https://doi.org/10.1037//0022-0663.93.3.521

[58] Naomi Yamashita and Toru Ishida. 2006. Effects of machine translation on collaborative work. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work – CSCW 2006*. ACM, New York, 515–524.

[59] Shlomit Yanisky-Ravid and Cynthia Martens. 2019. From the myth of Babel to Google Translate: Confronting malicious use of artificial intelligence: Copyright and algorithmic biases in online translations systems. *Seattle UL Rev.* 43 (2019), 99.

[60] Richard Zens, Franz Josef Och, and Hermann Ney. 2002. Phrase-based statistical machine translation. In *Annual Conference on Artificial Intelligence (LNCS 2479)*. Springer, Berlin, 18–32.