International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)
https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

Oral Reading Fluency Can Be Estimated Across Languages with Text-To-Speech Software

Alberto Luis Fernández[a*], Gabriel Enrique Jáuregui[b]

[a]*Facultad de Filosofía y Humanidades, Universidad Católica de Córdoba, Obispo Trejo 323, 5010, Córdoba, Argentina*
[b]*Facultad de Psicología, Universidad Nacional de Córdoba, Córdoba, Córdoba, Argentina*
[a]*Email: neuropsicologia.filo@ucc.edu.ar*
[b]*Email: gabriel.jaur@gmail.com*

Abstract

Oral reading fluency (ORF) is a good index of the reading skill level, measured as the number of words read correctly per minute (WRPM). However, ORF tests are not available in many languages. This study tested if the mean of WRPM could be estimated in languages for which ORF tests have not been developed by using free text-to-speech software. Mean time taken by Google Translate (GT) to read out loud 10 texts in 16 languages from the International Reading Speed Texts was compared with the mean time taken by human participants. An English/Other languages ratio was obtained for both reading systems. Both ratios were highly similar, showing that GT is a valid tool to estimate mean WRPM in multiple languages.

Keywords: Oral reading fluency; International Reading Speed Texts; text-to-speech software; cross-cultural assessment.

1. Introduction

Reading skills are fundamental to learning. It has been shown that poor readers struggle more at school. Reading ability has been tested with different parameters like accuracy, speed, prosody, or reading comprehension. Oral reading fluency (ORF), defined as "the oral translation of texts with speed and accuracy" [1], is one of those testing procedures. It has been deemed as a quick and easy informative testing technique. Asking a participant to read a short paragraph could be enough to estimate their ORF score. Research has shown that reading comprehension, learning, intelligence, school performance, and educational level [2, 3] are strongly related to ORF [1, 4, 5, 6]. Several studies have shown that ORF is a good predictor of reading comprehension at all levels, namely, elementary grades [7], secondary [8], and postsecondary levels [9]. Moreover, this relationship holds across languages. Similar results have been found in Korean [10], Norwegian [11].

Received: 9/6/2023
Accepted: 10/11/2023
Published: 10/23/2023
----

* Corresponding author.

Portuguese [12], Spanish [13], and Turkish [14]. Evidence from neuroimaging studies demonstrates that there are common cognitive mechanisms and neural networks for reading across languages, independently of the writing system. Readers of Spanish, English, Hebrew, and Chinese showed extensive convergence of printed and spoken language processing in many areas, including both cortical [bilateral inferior frontal gyrus, bilateral middle temporal gyrus to superior temporal gyrus, left inferior parietal lobule and subcortical (bilateral insula, putamen, thalamus) regions associated with both phonological and semantic processing [15, 16]. Furthermore, empirical evidence shows that there are reading universals and data shows that dyslexia manifests similarly across languages showing the same pattern of brain dysfunction [17].

ORF is therefore found at the base of a process: adequate ORF allows good reading comprehension which, in turn, allows adequate academic achievement. Academic achievement ultimately leads to better personal and social development and higher earnings [18, 19, 20, 21, 22].

The largest initiative to test ORF across languages is the International Reading Speed Texts

(IReST) which tests ORF in 17 languages [23]. The IReST consists of 10 texts for which normative data of words read per minute in each one of these languages have been developed. However, they represent only a small fraction of the more than 7,000 existing languages [24]. This situation marginalizes a huge population in the world that does not speak languages for which ORF tests have been developed. Unfortunately, developing ORF tests for many languages is not feasible mainly because of the lack of resources and the great variety of languages in many cultural environments. For example, there are 270 languages in India, which is a lower middle-income country according to the World Bank [25]. Out of these 270 languages, there are reading fluency tests only for English and Hindi [26]. Moreover, the test in Hindi is only for participants with up to three years of education. In South Africa, there are 11 official languages. Xhosa is one of the most spoken languages in the country and it is taught at schools in South Africa. However, these authors were unable to find any normative data on reading fluency in that language. In some countries, there is only one official language spoken by most people (e.g. Romania, Slovakia or Latvia). Thus, if there is no reading fluency normative data in that language, then the entire population of that country cannot be assessed.

Therefore, finding a method by which normative data for ORF tests in multiple languages can be derived from existing tests with normative data could have a significant impact on the educational field. This study shows that the mean of words read per minute can be estimated in languages for which ORF tests have not been developed by using free text-to-speech (TTS) software. This methodology is validated by comparing the performance of a text-to-speech software with existing ORF empirical data.

2. Methods

3. Materials

Ten texts from the IReST for the following 16 languages were included in this study: Arabic, Dutch, English, Finnish, French, German, Greek, Hebrew, Italian, Japanese, Turkish, Polish, Portuguese, Russian, Spanish, and Swedish.

4. Procedure

First, time taken to read each text by Google Translate (GT) was recorded. Each text was copied into Google Translate, then the software was asked to read each text translation aloud. Second, English was used as the reference or anchor language because most ORF studies were developed in English and English is one of the languages in which the number of words read per minute is higher. A ratio of words read per minute in English/Other languages was calculated for each one of the ten texts, using means provided by Trauzettel-Klosinski, Klaus Dietz and his colleagues (Trauzettel-Klosinski & Dietz, 2012). That investigation utilized newly created texts to evaluate the reading speed of a group of 436 individuals with normal vision who were native speakers of their respective languages. Each language had 25 participants, except for Japanese which had 36 participants. None of the texts were familiar to the readers. The participants, aged between 18 and 35, underwent an ophthalmologic/optometric examination to confirm normal or corrected-to-normal vision. Then, in this research, a mean of the ten ratios was obtained. This mean ratio allowed for the calculation of how many words per minute can be read in a specific language as compared to English. For example, the English/Portuguese mean ratio was 1.2, which means that for every word read in Portuguese, English readers can read 1.2 words. Next, a ratio of words read per minute in English/Other languages was calculated for each one of the ten texts using the means provided by the GT readings. Then, a mean of the ten ratios was obtained.

5. Results

Figure 1 shows the comparison between the mean time taken to read the ten texts in each language by participants in the Trauzettel-Klosinski, Klaus Dietz And his colleagues study and the mean time taken to read the ten texts in each language by GT. The correlation between both measures is r= .78. The most important differences were in Portuguese, Russian, and Hebrew in which GT read below the expected rate and Japanese, French, and Dutch in which it read at a higher speed than expected.

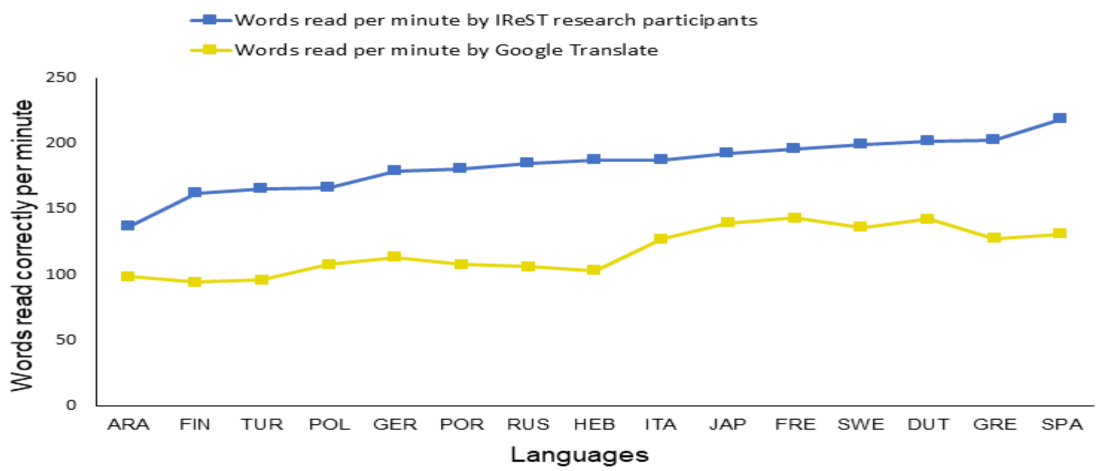

Figure 1: Mean time taken to read the ten IReST texts in each language by participants in the Trauzettel-Klosinski And his colleagues study and the mean time taken to read the texts in each language by Google Translate.

Figure 2 displays the ratios obtained between the mean time taken to read the ten texts in each language by human participants/GT. The ratios are all within the 1-2 range, which means that for every word read by human participants, GT reads between 1.36 (French) and 1.82 words (Hebrew).

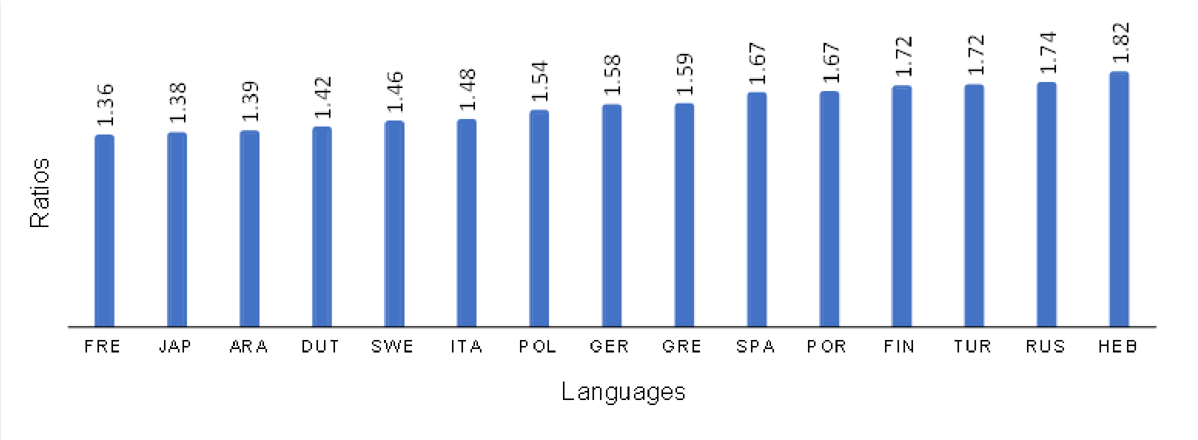

Figure 2: Ratios between the mean time taken to read the ten texts in each language by human participants / Google Translate. Each ratio was obtained by dividing the mean time taken by participants / mean time taken by Google Translate.

Table 1: Comparison of ratios to English between IReST research participants and Google Translate.

Languages	Ratio to English from IReST participants	Ratio to English from Google Translate
ARA	1.67	1.58
GRE	1.14	1.22
FIN	1.41	1.65
POL	1.37	1.43
TUR	1.38	1.62
GER	1.27	1.37
POR	1.26	1.44
RUS	1.23	1.46
HEB	1.22	1.50
ITA	1.22	1.22
JAP	1.19	1.11
FRE	1.17	1.08
SWE	1.15	1.14
DUT	1.13	1.09
SPA	1.04	1.14

Note. This table shows the ratios English/Other language obtained with the Google Translate and the International Reading Speed Texts study human participants. These ratios express how many words are read in English for each word read in other languages, i.e. Google Translate will read 1.58 words in English for each word read in Arabic.

Table 1 shows the ratios to English obtained from the IReST and GT. As observed, the differences between the ratios obtained with both systems are small. In all cases, the difference is a matter of one-hundredth of a word. For example, the ratio obtained from the IReST for German is 1.27, whereas the ratio obtained for the same language from the GT is 1.37. Table 2 exhibits the difference in the ratios to English obtained with both systems (IReST and GT). That difference was multiplied by the mean obtained in each language by the participants in the IReST study (second column). The last column shows the standard deviations obtained in each language by the participants in the IReST study. This procedure allowed us to establish the difference in the number of words produced as a result of the use of the ratio obtained from the GT. This number is within one standard deviation of the values obtained by the participants in the IReST research in all the cases, except for the last three languages in the list: Turkish, Finnish, and Hebrew. However, in these languages, the product exceeds one standard deviation criterion only by 1, 7, and 5 words, respectively.

Table 2: Difference between ratios to English obtained with GT and IReST participants.

Language	Difference	Diference x mean	Standard Deviation
FRE	-0.09	-12	26
ARA	-0.09	-10	20
JAP	-0.08	-12	30
DUT	-0.04	-6	29
SWE	-0.01	-1	34
ITA	0.00	0	28
POL	0.06	8	23
GRE	0.07	15	24
GER	0.09	12	17
SPA	0.14	20	28
POR	0.17	23	29
RUS	0.23	28	32
TUR	0.23	**26**	**25**
FIN	0.24	**25**	**18**
HEB	0.28	**34**	**29**

Note. This table shows the difference between the ratio English/Other language obtained with the Google Translate and by human participants in the International Reading Speed Texts study. The third column shows the result of the multiplication of this difference by the mean obtained by human participants. The last column shows the standard deviations obtained by the human participants. These results show that the values in the third column are within one standard deviation in all the languages except for the last three (numbers in bold letter).

6. Conclusions

The data in this study show that obtaining a ratio of English/Other languages with GT is highly similar to the ratio obtained with human participants. The progression shown in Figure 1, which displays ratios between the mean of words read per minute by human participants and GT; the high correlation between means shown in Figure 2; and, ultimately, the differences in the ratios obtained with both systems demonstrated that GT is a valid tool for estimating the equivalence in reading fluency in a given text between English and other languages. Although English was the anchor language in this study, any other language could have been used as the anchor since the data shows that the reading fluency relationship that exists between human participants reading in different languages is highly similar to the reading fluency relationship of a TTS software program reading in different languages. The differences in the ratios existing between both systems are insignificant as shown in Table 2 since most of them produce a difference that would exist in human participants reading these texts. The few cases over one standard deviation are exceeded by very few words, which makes them insignificant.

The most important impact of this study is that ORF can be estimated in languages for which ORF tests do not currently exist. For example, normative data for Spanish could be derived from the normative data obtained for an ORF test developed in English. The text of the ORF test is entered in GT and the time taken to read in English and Spanish (translation of the original text) is recorded. Then, the words read per minute in each language are estimated with the following formula: (60 x (number of words of the text))/the time GT takes to read the text. Next, the ratio can be computed as (English WRPM by GT/Spanish WRPM by GT). If, for example, the resulting ratio was 1.14, it means that for every word read in Spanish, it is expected that 1.14 words are read in English. Thus, if the mean words read per minute obtained by the normative data participants in English was, for example, 236, then the estimated mean of WRPM for the Spanish version of that text would be 207 (236/1.14). The same procedure can be performed with any other language for which TTS is enabled in GT.

The ratios obtained with IReST and GT are alike therefore the error is insignificant. This is demonstrated in Table 2. This error is within one standard deviation of the values obtained by the participants in the IReST, which represents the normal variability in their reading. In the few cases in which the one standard deviation criterion was exceeded, these values were within 1.5 standard deviations, which is also considered within normal variability. In addition, the standard deviation in the target language can also be estimated with the same procedure described above by entering, in the last step, the standard deviation from the normative data instead of the mean.

GT was selected for this research because it is the most popular, sophisticated, and easiest-to-access TTS software. It is a free software program and, as such, any person with Internet access can utilize it. There are sufficient studies supporting its development accuracy [27, 28, 29] and its function in over 100 languages, although developers plan to expand it to 1,000 languages [30]. Although TTS is not enabled for many languages that GT can translate, it is still the most comprehensive TTS software with currently 30 languages enabled. Although it was beyond the scope of this study, other TTS software might work as well.

The impact of this new methodology is multiple. It can help in the development of multiple ORF tests in many languages in which currently there are no such tests. This, in turn, will allow early detection of reading difficulties which allows early intervention. Some authors advocate the use of ORF for multiple purposes in the educational field, namely, screening, diagnosis, progress-monitoring and outcome [31]. Using ORF as screening, for example, can aid teachers in promptly recognizing students who are probably progressing well toward achieving future success in overall reading proficiency, while also identifying those who might require additional support. This can help solve individual difficulties as well as group difficulties. Slow progress in ORF can result in changes in the instruction program if group results are poor. As mentioned before, ORF is one contributing element found at the base of the pyramid that contributes to personal success and social development.

The procedure is simple, inexpensive, and quick to implement. Many ORF tests can be derived from currently existing tests. The only requirement is an accurate translation, for which GT becomes a valuable tool.

Another benefit of using this method is the possibility of including the world regions in which the resources for education and science are scarce. Moreover, it can help in the inclusion of language minorities in developed countries. Educational and/or health-related assessments of participants in their first language constitute a step forward in social justice and inclusion of diverse populations.

In addition, ORF has been used in neuropsychology to estimate the educational level of test subjects [2]. Education is a powerful variable that influences neuropsychological test performance: Therefore, it is very important to determine the educational level of test subjects. Educational level can be defined as “the level of cognitive development attained as a consequence of the education received” [2]. Although educational level has often been estimated by counting the number of years spent in school, educational level is currently considered more complex than the number of years in this estimation [32, 33]. ORF is a simple, short, and powerful way to estimate educational level. The procedure described in this study would allow for easy estimation of educational level in multiple languages based on simple existing ORF tests. This is particularly important for neuropsychologists testing immigrants and refugees since information about their educational background may not be as easy to obtain and many may have lower education. Estimating the educational level in populations with lower education is critical as it varies significantly among them.

In summary, this study presents a method by which ORF tests can be developed in different languages derived from existing tests by using freely available TTS software. The data presented supports the accuracy and readiness of the methodology. The potential benefits of the application of this methodology are manifold.

Acknowledgments

The authors would like to thank Kiana Paclibon, Gazul Rotela Leite, Ignacio Bellanti, and Julieta Castro Brunello.

References

[1]. L. S. Fuchs, D. Fuchs, M. K. Hosp, and J. R. Jenkins, “Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis”, *Scientific Stud. Reading*, vol. 5, no. 3, pp. 239–256, Jul. 2001. https://doi.org/10.1207/s1532799xssr0503_3

[2]. A. L. Fernandez and G. J. Arriondo, “Reading fluency as a measure of educational level”, *Dement. & Neuropsychol.*, vol. 15, no. 3, pp. 361–365, Sep. 2021. https://doi.org/10.1590/1980-57642021dn15-030008

[3]. G. Mackensen and H. Stichler, “Untersuchungen der Lesegeschwindigkeit in Abhängigkeit vom Bildungsgrad”, *Albr. Von Graefes Arch. Für Ophthalmol. Ver. Mit Arch. Für Augenheilkd.*, vol. 166, no. 1, pp. 81–86, 1963. https://doi.org/10.1007/bf00716108

[4]. L. Bigozzi, C. Tarchi, L. Vagnoli, E. Valente, and G. Pinto, “Reading Fluency As a Predictor of School Outcomes across Grades 4–9”, *Frontiers Psychol.*, vol. 8, Feb. 2017. https://doi.org/10.3389/fpsyg.2017.00200

[5]. S. A. Stage and M. D. Jacobsen, “Predicting Student Success on a State-mandated Performance-based

Assessment Using Oral Reading Fluency", *School Psychol. Rev.*, vol. 30, no. 3, pp. 407–419, Sep. 2001. https://doi.org/10.1080/02796015.2001.12086123

[6]. R. P. Carver, "Intelligence and reading ability in grades 2–12", *Intelligence*, vol. 14, no. 4, pp. 449–455, Oct. 1990. https://doi.org/10.1016/s0160-2896(05)80014-5

[7]. Y.-S. Kim, R. K. Wagner, and E. Foster, "Relations Among Oral Reading Fluency, Silent Reading Fluency, and Reading Comprehension: A Latent Variable Study of First-Grade Readers", *Scientific Stud. Reading*, vol. 15, no. 4, pp. 338–362, Jul. 2011. https://doi.org/10.1080/10888438.2010.493964

[8]. T. V. Rasinski, N. D. Padak, C. A. McKeon, L. G. Wilfong, J. A. Friedauer, and P. Heim, "Is Reading Fluency a Key for Successful High School Reading?", *J. Adolescent & Adult Literacy*, vol. 49, no. 1, pp. 22–27, Sep. 2005. https://doi.org/10.1598/jaal.49.1.3

[9]. S. Seok and B. DaCosta, "Oral Reading Fluency as a Predictor of Silent Reading Fluency at Secondary and Postsecondary Levels", *J. Adolescent & Adult Literacy*, vol. 58, no. 2, pp. 157–166, Sep. 2014. https://doi.org/10.1002/jaal.342

[10]. Y.-S. Kim, C. H. Park, and R. K. Wagner, "Is oral/text reading fluency a "bridge" to reading comprehension?", *Read. Writ.*, vol. 27, no. 1, pp. 79–99, Mar. 2013. https://doi.org/10.1007/s11145-013-9434-7

[11]. A. Arnesen, J. Braeken, S. Baker, W. Meek-Hansen, T. Ogden, and M. Melby-Lervåg, "Growth in Oral Reading Fluency in a Semitransparent Orthography: Concurrent and Predictive Relations With Reading Proficiency in Norwegian, Grades 2-5", *Reading Res. Quart.*, vol. 52, no. 2, pp. 177–201, Aug. 2016. https://doi.org/10.1002/rrq.159

[12]. M. A. Martins and S. A. Capellini, "Relação entre fluência de leitura oral e compreensão de leitura", *CoDAS*, vol. 31, no. 1, 2019. https://doi.org/10.1590/2317-1782/20182018244

[13]. M. Álvarez-Cañizo, P. Suárez-Coalla, and F. Cuetos, "The Role of Reading Fluency in Children's Text Comprehension", *Frontiers Psychol.*, vol. 6, Nov. 2015. https://doi.org/10.3389/fpsyg.2015.01810

[14]. Yildirim, K., & Rasinski, T. Reading fluency beyond English: Investigations into reading fluency in Turkish elementary students. *International Electronic Journal of Elementary Education*, 7(1), 97-106. 2014.

[15]. J. G. Rueckl *et al.*, "Universal brain signature of proficient reading: Evidence from four contrasting languages", *Proc. Nat. Acad. Sci.*, vol. 112, no. 50, pp. 15510–15515, Nov. 2015. https://doi.org/10.1073/pnas.1509321112

[16]. L. Verhoeven and C. Perfetti, "Universals in Learning to Read Across Languages and Writing Systems", *Scientific Stud. Reading*, pp. 1–15, Jun. 2021. https://doi.org/10.1080/10888438.2021.1938575

[17]. F. Richlan, "The Functional Neuroanatomy of Developmental Dyslexia Across Languages and Writing Systems", *Frontiers Psychol.*, vol. 11, Feb. 2020. https://doi.org/10.3389/fpsyg.2020.00155

[18]. J. S. Zax and D. I. Rees, "IQ, Academic Performance, Environment, and Earnings", *Rev. Econ. Statist.*, vol. 84, no. 4, pp. 600–616, Nov. 2002. https://doi.org/10.1162/003465302760556440

[19]. A. Z. Curi and N. Menezes-Filho, "The relationship between school performance and future wages in Brazil", *EconomiA*, vol. 15, no. 3, pp. 261–274, Sep. 2014. https://doi.org/10.1016/j.econ.2014.08.001

[20]. Olneck, M. Economic Consequences of the Academic Achievement Gap for African Americans.

Marquette Law Review, 89, 95-116. 2005

[21]. J. Savage, C. J. Ferguson, and L. Flores, “The effect of academic achievement on aggression and violent behavior: A meta-analysis”, *Aggression Violent Behav.*, vol. 37, pp. 91–101, Nov. 2017. https://doi.org/10.1016/j.avb.2017.08.002

[22]. P. J. Hines, B. Wible, and M. McCartney, “Learning to Read, Reading to Learn”, *Science*, vol. 328, no. 5977, p. 447, Apr. 2010. https://doi.org/10.1126/science.328.5977.447

[23]. S. Trauzettel-Klosinski, K. Dietz, y IReST Study Group, “Standardized assessment of reading performance: the New International Reading Speed Texts IReST”, *Invest. Ophthalmol. Vis. Sci.*, vol. 53, núm. 9, pp. 5452–5461, 2012.

[24]. Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2023. Ethnologue: Languages of the World. Twenty-sixth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

[25]. “World Bank Open Data,” World Bank Open Data. [Online]. Available: https://data.worldbank.org/country/IN. [Accessed: 26-Jul-2023].

[26]. S. Vagh, “Validating the ASER testing tool : Comparisons with reading fluency measures and the read India measures,” 2012.

[27]. Y. Wu et al., “Google’s Neural Machine Translation system: Bridging the gap between human and Machine Translation,” *arXiv* [cs.CL], 2016.

[28]. E. C. Khoong, E. Steinbrook, C. Brown, and A. Fernandez, “Assessing the use of Google translate for Spanish and Chinese translations of emergency department discharge instructions,” *JAMA Intern. Med*., vol. 179, no. 4, pp. 580–582, 2019.

[29]. J. L. Jackson et al., “The accuracy of Google translate for abstracting data from non–English-language trials for systematic reviews,” *Ann. Intern. Med*., vol. 171, no. 9, p. 677, 2019.A.

[30]. Bapna et al., “Building machine translation systems for the next thousand languages,” *arXiv* [cs.CL], 2022.

[31]. J. Hasbrouck and G. A. Tindal, “Oral reading fluency norms: A valuable assessment tool for reading teachers,” *Read. Teach*., vol. 59, no. 7, pp. 636–644, 2006.

[32]. A. L. Chin, S. Negash, S. Xie, S. E. Arnold, and R. Hamilton, “Quality, and not just quantity, of education accounts for differences in psychometric performance between african americans and white non-hispanics with Alzheimer’s disease,” *J. Int. Neuropsychol*. Soc., vol. 18, no. 2, pp. 277–285, 2012.

[33]. P. Sayegh, A. Arentoft, N. S. Thaler, A. C. Dean, and A. D. Thames, “Quality of education predicts performance on the Wide Range Achievement Test-4th Edition Word Reading subtest,” *Arch. Clin. Neuropsychol.,* vol. 29, no. 8, pp. 731–736, 2014.