

# Examining the Quality of Machine Translation

*Derek Acosta*

## Analyze the Translations

Go over at least 20 sentences, manually correct each sentence, and report for each sentence:

1. the source sentence
2. the statistical machine translation
3. the neural machine translation
4. a correction of the machine translation
5. an assessment of the error(s) in the statistical machine translation
6. an assessment of the error(s) in the neural machine translation
7. whether you picked the statistical or neural translation as the better one

You may do steps 5. and 6. in any way you want. For instance, you could classify errors as “reordering errors”, “word sense error for a noun”, or any other type of error you can think of.

**My Language: Spanish**

**Site:** <http://cnnespanol.cnn.com/2018/01/25/con-una-tasa-de-analfabetismo-de-27-es-bolivia-un-ejemplo-en-educacion-para-america-latina/>

### sentence 1

1. Morales citó cifras del Banco Mundial para hablar sobre analfabetismo y esperanza de vida, que según desde 1995 a la fecha han tenido una reducción importante en su país.
2. Morales cited figures from the World Bank to discuss illiteracy and life expectancy, as from 1995 to date they have had a significant reduction in their country.
3. Morales cited World Bank figures to talk about illiteracy and life expectancy, which from 1995 to date have had a significant reduction in his country.
4. Morales cited that the Wolrd Bank figures to talk about illiteracy and life expectancy, which from 1995 to date have had a significant reduction in his country.
5. **Error assessment of the statistical machine translation** word choice is a little awkward but still optimal
6. **Error assessment of the neural machine translation** Nothing egregiously wrong with this sentence but uses wording that is not the best
7. statistical

### sentence 2

1. El Banco Mundial no reporta cifras de alfabetización en 1995 para Bolivia, sino en 1992, y los datos más recientes corresponden a 2015.
2. The World Bank does not report figures for literacy in 1995 for Bolivia, but in 1992, and the latest data correspond to 2015.
3. The World Bank does not report literacy figures in 1995 for Bolivia, but in 1992, and the most recent data correspond to 2015.

4. The World Bank does not report Bolivia's literacy reports for 1995, if not and as well as 1992, and the most recent data corresponds to 2015.
5. **Error assessment of the statistical machine translation** slight issues involving making 'correspond' singular  
sentence structure appears to be a little off.  
uses a lot of passive voice that makes the sentence read a bit oddly
6. **Error assessment of the neural machine translation** slight issues involving making 'correspond' singular  
sentence structure appears to be a little off.  
uses a lot of passive voice that makes the sentence read a bit oddly
7. statistical

### sentence 3

1. La reducción en casi 20 años fue de 12,4 puntos porcentuales.
2. The reduction in almost 20 years was 12.4 percentage points.
3. The reduction in almost 20 years was 12.4 percentage points.
4. The reduction in almost 20 years was 12.4 percentage points.
5. **Error assessment of the statistical machine translation** The word choice used was that, which might explain that the two sentences are identical.
6. **Error assessment of the statistical machine translation** The word choice used was that, which might explain that the two sentences are identical.
7. Both work

### sentence 4

1. Y, en 1992, la tasa de analfabetismo que reportó Bolivia fue de 19,69%.
2. And, in 1992, the rate of illiteracy which reported Bolivia was 19.69%.
3. And, in 1992, the illiteracy rate reported by Bolivia was 19.69%.
4. And, in 1992, Bolivia reported that the rate of illiteracy was 19.69%
5. **Error assessment of the statistical machine translation** when choosing between the turn of phrase 'rate of illiteracy' and 'illiteracy rate' this is truly up to personal preference whether someone would like to use one over the other. In this case, the phrase do not really provide any more information causing ambiguity.
6. **Error assessment of the neural machine translation** See above.
7. Either works

### sentence 5

1. En 2012, el índice general de analfabetismo para Bolivia era del 5,5%.
2. In 2012, the general illiteracy rate for Bolivia was 5.5%.
3. In 2012, the general illiteracy rate for Bolivia was 5.5%.
4. **In 2012, the general illiteracy rate for Bolivia was 5.5%.** The word choice used was that, which might explain that the two sentences are identical.

5. **Error assessment of the statistical machine translation** The word choice used was that, which might explain that the two sentences are identical.
6. Both work

#### sentence 6

1. 2017 cerró con una tasa de analfabetismo de 2,52%, según le dijo a CNN en Español Ramiro Tolaba, director del Programa de Alfabetización y Post Alfabetización del Ministerio de Educación.
2. 2017 closed with a 2.52% illiteracy rate, as he told CNN in Spanish Ramiro Tolaba, director of the program of literacy and Post literacy of the Ministry of education.
3. 2017 closed with an illiteracy rate of 2.52%, according to CNN in Spanish, Ramiro Tolaa, director of the literacy and Post-literacy Program of the Ministry of Education.
4. 2017 closed with a illiteracy rate of 2.52%, as he told CNN in Spanish Ramiro Tolaba, director of the program for literacy and Post-literacy of the Ministry of Education.
5. **Error assessment of the statistical machine translation** The capitalization of proper nouns is lacking, in comparison to the neural.  
also lacks hyphening usage
6. **Error assessment of the neural machine translation** Nothing incredible wrong with this sentence - it passes.
7. neural

#### sentence 7

1. Según el Gobierno de Bolivia, el éxito se lo debe la implementación de un Programa Nacional de Alfabetización cuando Evo Morales llegó al poder en 2006.
2. According to the Government of Bolivia, the success implementation of a national literacy program is due when Evo Morales came to power in 2006.
3. According to the Bolivian government, success is due to the implementation of a national literacy program when Evo Morales came to power in 2006.
4. According to the Government of Bolivia, the successful implementation of a national literacy program is due to when Evo Morales came into power in 2006.
5. **Error assessment of the statistical machine translation** requires a higher use of determiners  
the term success implementation is incorrect and should be wrong to the adjective 'successful'
6. **Error assessment of the neural machine translation** the term success implementation is incorrect and should be wrong to the adjective 'successful'  
Otherwise this sentence works well.
7. neural

#### sentence 8

1. "Cuando él entra al gobierno, encuentra el analfabetismo como uno de los temas más álgidos del país y lo plantea como una de las acciones que había que resolver", le dijo Tolaba a CNN en Español.

2. “When it comes to the Government, is illiteracy as one of the most high topics of the country and poses as one of the actions that had to be solved” said Tolaba CNN in Spanish.
3. “When he enters the government, he finds illiteracy as one of the most highlights issues in the country and raises it as one of the actions that had to be resolved,” Tolan told CNN in Spanish.
4. “When he enters the government, he finds illiteracy to be one of the most highlighted issues in the country and still raises it as one of the actions that had to be resolved,” Tolan told CNN in Spanish.
5. **Error assessment of the statistical machine translation** ‘high topics’ sounds a bit awkward and should use, as the neural sentence did, the phrase ‘most highlighted’ to convey the sentence with more accuracy.
6. **Error assessment of the neural machine translation** Does use past tense as well. Should change highlights to highlighted.
7. neural

#### sentence 9

1. Según él, la implementación del programa ‘Yo Sí Puedo’ permitió que entre 2006 y 2008 las tasas de analfabetismo en Bolivia pasaran del 13,28% al 4,7%.
2. According to him, the implementation of the program ‘Yes I’ allowed that between 2006 and 2008 rates of illiteracy in Bolivia will 13.28% to 4.7%.
3. According to him, the implementation of the ‘I can do’ programme allowed between 2006 and 2008 illiteracy rates in Bolivia to go from 13.28% to 4.7%.
4. According to him, the implementation of the program ‘Yo Sí Puedo’ allowed that between 2006 and 2008 the illiteracy rates in Bolivia went from 13.28% to 4.7%.
5. **Error assessment of the statistical machine translation** The program’s name should have been translate, or if it has it should have provide the Spanish learning because it a proper noun.  
not as explicit with language usage.
6. **Error assessment of the neural machine translation** The program’s name should have been translate, or if it has it should have provide the Spanish learning because it a proper noun.
7. neural

#### sentence 10

1. El programa ‘Yo Sí Puedo’ tiene sus raíces la Cuba de Fidel Castro y fue elaborado en el año 2000 por la pedagoga Leonela Relys.
2. ‘Yes I’ program has its roots Fidel Castro’s Cuba and was drafted in 2000 by the pedagogue Leonela Relys.
3. The ‘I can do’ program is rooted in Fidel Castro’s Cuba and was developed in the year 2000 by the pedagogue Leonela Relys.
4. The program ‘Yo Sí Puedo’ has its roots in Fidel Castro’s Cuba and was developed in 2000 by the pedagogue Leonela Relys.
5. **Error assessment of the statistical machine translation** use the term drafted to explain for when something was established

once again translates a proper noun for consideration of context

6. **Error assessment of the neural machine translation** once again translates a proper noun for consideration of context
7. statistical

#### sentence 11

1. Son 17 cintas previamente grabadas en Cuba, con 65 lecciones que para 2006 ya estaban presentes en 15 países para alfabetizar a miles de personas, según la Unesco.
2. They are 17 tapes previously recorded in Cuba, with 65 lessons that, for 2006, were already present in 15 countries for literacy to thousands of people, according to Unesco.
3. There are 17 tapes previously recorded in Cuba, with 65 lessons that for 2006 were already present in 15 countries to literacy thousands of people, according to UNESCO.
4. There are 17 tapes previously recorded in Cuba, with 65 lessons that by 2006 were already present in 15 countries to teach literacy to thousands of people, according to Unesco.
5. **Error assessment of the statistical machine translation** odd comma usage
6. **Error assessment of the neural machine translation** Made an incorrect acronym usage and capitalized it.
7. neural

#### sentence 12

1. “El Programa Nacional de Alfabetización empezó con una metodología cubana que fue contextualizada a la realidad boliviana”, dijo Tolaba a CNN en Español.
2. “The national literacy program started with a Cuban methodology that was contextualized to the Bolivian reality”, Tolaba told CNN in Spanish.
3. “The National Literacy program began with a Cuban methodology that was contextualized to the Bolivian reality,” Tolaa told CNN in Spanish.
4. “The National Literacy Program began with a Cuban methodology that was contextualized to the Bolivian reality,” Tolaba told CNN en Español.
5. **Error assessment of the statistical machine translation** incorrect proper noun capitalization.
6. **Error assessment of the neural machine translation** Nothing particularly wrong with this sentence
7. neural

#### sentence 13

1. El objetivo es que las personas aprendieran a leer, escribir, y aprendan aritmética básica, en siete semanas a través de videocasetes, recibiendo los cursos de lunes a viernes dos horas al día.
2. The goal is for people to learn to read, write, and learn basic arithmetic in seven weeks through video cassettes, receiving courses from Monday to Friday two hours a day.
3. The goal is for people to learn to read, write, and learn basic arithmetic, in seven weeks through videocassettes, receiving courses from Monday to Friday two hours a day.

4. The goal is for people to learn to read, write, and learn basic arithmetic, in seven weeks through videocassettes, receiving the courses from Monday to Friday two hours a day.
5. **Error assessment of the statistical machine translation** the problem here lies in whether to use the oxford comma but no rule are broken and is clear to understand
6. **Error assessment of the neural machine translation** the problem here lies in whether to use the oxford comma but no rule are broken and is clear to understand
7. Either works

#### sentence 14

1. Durante 11 años de programa se alfabetizó a más de un millón de personas (en un país de más de 11 millones de habitantes) a través de sus programas de alfabetización y posalfabetización (Yo Sí Puedo y Yo Sí Puedo Seguir, respectivamente).
2. Over 11 years of program is alfabetizó more than one million people (in a country of more than 11 million inhabitants) through its literacy programs and posalfabetizacion (yes I and I yes I can follow, respectively).
3. Over 11 years of program, more than one million people (in a country of more than 11 million inhabitants) were literated through their literacy and postliteracy programs (I can and I can follow, respectively).
4. During 11 years of the program, more than one million people (in a country of more than 11 million inhabitants) were literate through their literacy and post-literacy programs (Yo Sí Puedo and Yo Sí Puedo Seguir, respectively).
5. **Error assessment of the statistical machine translation** Does translate a word and sticks with using the Spanish words such as ‘alfabetizó’ and ‘posalfabetizacion’ once again translate proper nouns
6. **Error assessment of the neural machine translation** more explicit with its language usage and commits to use the correctly translated Spanish term
7. neural

#### sentence 15

1. En un informe de 2006, la Unesco resaltó los beneficios de este programa por su “enfoque conceptual de la alfabetización, el aprendizaje, las competencias para la vida diaria y la movilización social” y le otorgó el Premio de Alfabetización “Rey Sejong” ese año.
2. In a 2006 report, Unesco stressed the benefits of this program for their “conceptual focus on literacy, learning, skills for daily life and social mobilization” and awarded him the prize for literacy “King Sejong” that year.
3. In a 2006 report, UNESCO highlighted the benefits of this program for its “conceptual approach to literacy, learning, competencies for daily life and social mobilization” and awarded it the “Rey-Jong” literacy award that year.
4. In a 2006 report, Unesco highlighted the benefits of this program for its “conceptual approach to literacy, learning, life skills and social mobilization” and awarded it the “King Sejong” Literacy Prize that year .
5. **Error assessment of the statistical machine translation** Good translation nothing wrong to my eyes.

6. **Error assessment of the neural machine translation** Incorrect acronym capitization usage  
awkward tense and word association use.
7. statistical

#### sentence 16

1. Tolaba dice que los programas de alfabetización y posalfabetización han sido implementados en lenguas indígenas bolivianas como aimara, quechua y guaraní y que muchos de los beneficiados han seguido los estudios de primaria, secundaria y hasta estudios de nivel superior.
2. Tolaba says that literacy and posalfabetizacion programs have been implemented in Bolivia as Aymara, quechua and guarani indigenous languages and many of the winners have followed studies of primary, secondary and even level studies upper.
3. Tolan says that the literacy and post-literacy programs have been implemented in Bolivian indigenous languages such as Aimara, Quechua and Guarani and that many of the beneficiaries have followed the primary, secondary and even level studies Top.
4. Tolaba says that literacy and post-literacy programs have been implemented in indigenous Bolivian languages such as Aymara, Quechua and Guarani and that many of the beneficiaries have continued their studies in primary, secondary and even higher education.
5. **Error assessment of the statistical machine translation** Use Spanish dictation and does not do a proper translation
6. **Error assessment of the neural machine translation** Its a run on sentence but other than this it works well in conveying the message
7. neural

#### sentence 17

1. Y justamente esa educación “intra e intercultural” en el lenguaje materno de los indígenas es lo que resalta Adán Pari, experto en educación de Unicef en Bolivia.
2. And precisely that education “intra and intercultural” in the maternal language of the indigenous people is what stands out Adán Pari, expert in education of Unicef in Bolivia.
3. And just that “intra-and intercultural” education in native language is what highlights Adam Pari, a UNICEF education expert in Bolivia.
4. And it is precisely this “intra and intercultural” education in the native language of the indigenous that is highlighted by Adán Pari, an education expert at Unicef in Bolivia.
5. **Error assessment of the statistical machine translation** transitation phrase does not work well and does not provide much content for the expert.
6. **Error assessment of the neural machine translation** odd hypen use but other than that one issue the sentence works very well.
7. neural

#### sentence 18

1. “Es importante considerar que Bolivia siendo un país con un alto porcentaje de pueblos indígenas considera necesariamente la educación intra e intercultural y plurilingüe”, le dijo Pari a CNN en Español.

2. “It is important to consider that Bolivia is a country with a high percentage of indigenous peoples considered necessarily intra education and intercultural and plurilingual ’, said Pari to CNN in Spanish.
3. “It is important to consider that Bolivia being a country with a high percentage of indigenous peoples necessarily considers intra and intercultural and multilingual education,” Pari told CNN in Spanish.
4. “It is important to consider that Bolivia being a country with a high percentage of indigenous peoples necessarily considers intra and intercultural and multilingual education,” Pari told CNN en Español.
5. **Error assessment of the statistical machine translation** uses an incorrect translated word ‘plurilingual’ when I believe what they intended ‘multilingual’  
odd transitation phrase
6. **Error assessment of the neural machine translation** Works very well
7. neural

#### sentence 19

1. Pero en un informe de 2017 sobre Conocimiento Indígena y Políticas Educativas Educativas en América Latina, la Unesco dice que aunque a nivel nacional la tasa de alfabetismo sea de 92,3%, esta se reduce a 81,07% entre la población cuyo idioma materno es indígena.
2. But in a report by 2017 on indigenous knowledge and education educational policy in Latin America, Unesco said that although nationwide literacy rate of 92.3%, this is reduced to 81,07% among the population whose mother tongue is indigenous.
3. But in a report of 2017 on indigenous knowledge and educational education policies in Latin America, UNESCO says that although at the national level the literacy rate is 92.3%, this is reduced to 81.07% among the population whose native language is indigenous.
4. But in a 2017 report on Indigenous Knowledge and Educational Educational Policies in Latin America, Unesco says that although at the national level the literacy rate is 92.3%, it is reduced to 81.07% among the population whose mother tongue It is indigenous.
5. **Error assessment of the statistical machine translation** the translation of the phrase “education educational policy” does not work well, which is sometihng that neural got right.
6. **Error assessment of the neural machine translation** Incorrect proper noun use but the rest works well.
7. neural

#### sentence 20

1. Los críticos de este programa dicen que más que las cifras, lo que debe analizarse es la calidad de la educación.
2. Critics of the program say that rather than you the figures, to be analysed is the quality of education.
3. Critics of this program say that more than the figures, what needs to be analyzed is the quality of education.
4. Critics of this program say that more than figures, what should be analyzed is the quality of education.



5. **Error assessment of the statistical machine translation** word order choice is strange.  
the use of the comparison operates does not translate well. 'Rather than' should become  
'more than' which is something the neural suggests
6. **Error assessment of the neural machine translation** well constructed.
7. neural