# Variation Impoverishment resulting from Machine Translations: Empirical Evidence from Spanish Counterfactual Predicates

Isabel Repiso

# Variation impoverishment resulting from machine translations: Empirical evidence from Spanish counterfactual predicates

**Dr. Isabel Repiso**
University of Salzburg, Department of Romance Languages
Erzabt-Klotz-Straße 1, 5020 Salzburg (Austria)

**Abstract**:

*The present article compares online machine translations and human translations in the aim of describing variation effects in Spanish output segments. Our survey focuses in a counterfactual construction existing both in English and Spanish but whose uses and frequencies differ in each language. English perfective Should have predicates have a formal equivalent in Spanish –i.e. Debería haber + past participle– although the functional equivalent used by Spanish-speakers authors is a free-modal subjunctive construction –i.e. Hubiera + past participle (Repiso 2018). Our survey is based on 1.7 million-word Social Sciences corpus covering 8 essays, 4 political biographies and 2 dystopian novels. In all, 95 perfective Should have predicates were elicited in English. The human and machine translations in Spanish were subsequently analyzed. Our results show that human translations significantly preserve a wider range of lexical and morphological features compared to machine translations, which in turn favor prominently the formal equivalent construction Debería haber. The effect of this is the anglizing of the target language's morphosyntactic level by a quantitative reversal in the constructions pertaining to the irrealis semantic domain. Our survey suggests that a systemic use of MT risks to impoverish the target language outputs and thus, to oversimplify human communication. In a broader sense, MT challenges two phenomena widely observed in linguistics –i.e., natural languages' richness and individual differences within groups.*

**Keywords**: machine translation, language variation, form and content mapping, transfer.

## Introduction

Because variation is a transversal component of human groups' behavior, it has been at the center of the research agenda of many disciplines within and beyond social sciences. In the domain of linguistics, language variation has been explored as an intriguing subject essential to any understanding of human communication (Glowka & Lance 1993). In particular, it has been approached by dialectologists, sociolinguists, scholars interested in language universals and language acquisition[1]. The present study should contribute to shed light upon the effects of online machine translations compared to human translations. Language variation can occur across language users and also within individual grammars. A single speaker will make systematic choices in pronunciation, morphology, word choice, and grammar reflecting a range of non-linguistic factors (Conrad & Biber 2001). In a similar way, translators are led to make choices in the target language at the lexical level between competitor terms and at the syntactic level between concurrent constructions. If one assumes that register variation is all-pervasive in human language, then the question arises whether machines will preserve the same degree of variation in a translation task. Testing this point is the purpose of our survey. Our article is structured in five sections. In the first section, we highlight some interesting results of some recent studies dealing with machine translations. We subsequently introduce the semantic domain of counterfactuality to which the English perfective *Should have* is frequently associated. The Spanish constructions pertaining to the *irrealis* domain are also discussed in the aim of distinguishing the formal equivalent construction from the functional one. In the third section we present the corpus of written texts and translations that we have reunited, as well as the methods used (i) to elicit the utterances containing *Should have* and (ii) to run statistical analyses. The Spanish translations performed by humans and machines are presented in the 4rd section. Finally we discuss our findings and relate them to previous studies. Conclusions are drawn in the final section.

---

[1] Cf. Siemund (2011) for an overview of generativists studies dealing with language variation, and Bayley & Preston (1996) for second language acquisition and linguistic variation.

## 1. Previous studies dealing with machine translations

To our knowledge, the impact of MT in language variation is a question relatively unexplored in the field of translation and corpus-based studies[2]. The comparison between human and machine translations (MT) has been at the center of some empirical surveys interested in the accuracy of specialized terminology (Schulz et al. 2013), and in productivity gains (Toral et al. 2018). In what concerns productivity, it has been shown that both statistical and neural machine translation have positive effects in post-editing, increasing the production of words per hour by 18% and 36% respectively, compared to scratch (Toral et al. 2018). Regarding medical terminology, human translations proved to be rated significantly higher over machine translations as for content fidelity and linguistic correctness (Schulz et al. 2013). Machine translations have also been approached lonely as a challenging field regarding the restitution of proverbs (Al-khresheh & Almaaytah 2018) and the outputs' quality (Lotz & Van Rensburg 2016). In a longitudinal survey over four years, quality improvements were found in online machine outputs in errors related to mistranslations and syntax components (Lotz & Van Rensburg 2016). However, MT still results in literal translation when confronted to proverbs in the Arabic-English language configuration (Al-khresheh & Almaaytah 2018).

## 2. The perfective Should have and its equivalents in Spanish

The present study investigates the translation of the English perfective construction *Should have* + Past participle, which is generally associated to counterfactual meanings. Pietrandrea (2010) defines counterfactuality as a semantic subdomain of irreality. From a semantic point of view, counterfactuality combines the apparently contradictory features of potentiality and non-actualization (Versatrete 2005). Counterfactual sentences express something that was desired but did not happen, or an action that was intended but not carried out in the end, like in the example below.

(1) The police should have done something to prevent the killing (Van linden & Verstraete, 2008: 1866)

The speaker's state-of-knowledge at the moment of the topic time plays a crucial role in processing a meaning as potential or counterfactual. Counterfactuality implies the speaker's certainty about the non-actualization of the propositional content, whereas potentiality implies the speaker's uncertainty about the actualization of the propositional content. One is more likely to be certain about present or past states-of-affairs rather than future states-of-affairs. This is why some authors insist to place counterfactuality in the non-future time (Iatridou 2000, Vetters 1994). The temporal span to which counterfactual sentences refer to is actually a future within the past (Grevisse & Goosse 2008). Thus, in the example above the construction *Should have* + Past participle expresses a temporal value of non-factual posteriority rooted in the past.

Past tense morphology has been mentioned as a critical ingredient of counterfactual morphology (Iatridou 2000), all along with modal markers (Van linden & Verstraete 2008). Typological differences between Spanish and French have been observed in the oral production of counterfactual scenarios by native speakers (Repiso 2014). In a mutation task consisting of providing alternative scenarios to factual states-of-affaires native Spanish speakers significantly used the perfect subjunctive with no modal markers (e.g., que la policía hubiera intervenido, *that the police had intervened*), whereas native French speakers privileged the combination of the past conditional and a modal marker (e.g., la police aurait dû intervener, *police should have intervened*). The frequencies of these grammatical devices proved to be a salient difference between French and Spaniards' productions.

In Spanish, counterfactual readings can be conveyed by different tenses and moods. According to grammarians, the most frequent tenses are the pluperfect subjunctive (cf. examples 2 & 3), the past infinitive within iterative sentences (4), and the past conditional within *if-* clauses' main sentences (5). From a semantic viewpoint, example (2) usually denotes the speaker's dismissing about the topic or theme conversation. The pluperfect subjunctive is used in a concessive manner to reject an eventuality that the speaker considers false. Within negative sentences (3), the pluperfect subjunctive denotes a polarity and, at the same time, asserts the lexical content of the verbal segment.

(2) Me hubieran avisado (RAE 2009: 3136) (In that case) they had warned me
(3) No te hubieras demorado tanto (RAE 2009: 3136) You better ain't be so late

---

[2] The starting point in corpus-based studies is a description of authentic examples of language use. Traditionally, corpora have been used diachronically or synchronically to explore and describe how languages are actually used in written and oral modality. In this context, language variation has emerged as an interesting subject worthily of explanation. The findings in this field are thus related to systematically-collected samples of authentic language production and to variation, which is intended as different preferences for use under different conditions (Yamazaki & Sigley 2013).

(4) ¡Haber venido ayer! (RAE 2009: 3572)  You should have come yesterday

(5) Si se lo hubieran explicado, lo habría entendido (RAE 2009: 3573) If they had explained it to her/him, she/he would have understood

If they had explained it to her/him, she/he would have understood

Beyond flexional features like tenses the notion of probability can also be encoded in Spanish by lexical means, like the periphrastic segments *Tener que* + Infinitive and *Deber* + Infinitive. The former is more frequent than the latter and conveys a modal message that is presented by the speaker as something inevitable (RAE 2009: 2144). Unlikely, the periphrastic *Deber* + Infinitive conveys a message socially desirable or accordingly to the norms of a given community. Both of these formulae are used to denote epistemic values in a wide range of indicative tenses such as the present indicative (6), the Spanish imperfecto (7), the past simple (8) and the present conditional (9). Thus, the periphrases *Tener que* or *Deber* + Infinitive can be seen as different points within a scale of intensity or evaluative language in Spanish. This raises the question whether the construction *Should have* is more frequently used in English with a binding connotation or a desirable connotation. It may well be that in some contexts the semantic equivalence of *Should have* is the Spanish *Tener que*, whereas in other contexts *Deber que* seems more accurate[3]. In any case, the translator seems to have some degree of choice at this level.

(6)  Son muy listos y además tienen que haber hecho estudios (RAE 2009: 2145) They are very smart, plus they might have completed some degree

(7)  Escobedo debía haber llegado a Madrid, pero no había noticias (RAE 2009: 2145) Escobedo should have arrived at Madrid though there was no confirmation

(8)  No me extraña que se matara […], debí haberlo previsto (RAE 2009: 2145) Her/his death didn't surprise me […], I should have expected it

(9)  Tendría que encender las luces […] y gritar de dolor (RAE 2009: 2144) I/she/he would turn the lights on […] and scream out with pain

Among the idiomatic constructions provided by grammarians, the pluperfect subjunctive of examples (2) & (3) proved to be prominent in the Spanish reference corpus CREA[4], compared to the periphrastic construction *Deber* + infinitive of examples (7) & (8). In other words, the former is more frequently used by Spanish-speaking authors, journalists and writers, compared to the latter (Repiso 2018: 151).

## 3. Materials and methods

Our study is based on 1.7 million-word Social Sciences corpus covering 8 essays, 4 political biographies and 2 dystopian novels. Our purpose was to collect the same sample of texts both in English and Spanish, so that we could work on a parallel corpus. We reunited texts from political activists Naomi Klein and Noam Chomsky (2 each), political leaders Nelson Mandela, Barack Obama and Malala Yousafzai, psychologist Daniel Kahneman and feminists Audre Lorde, Silvia Federici and Judith Butler. We also included Virginia Woolf's essay *A room of one's own*, Ray Bradbury's *Fahrenheit 451* and George Orwell's *1984*. The nature of these 14 texts is diverse. Among the essays, Kahneman's is devoted to the cognitive mechanisms guiding decision-making processes. The rest of them are critical with hegemonic representations such as capitalism –Klein's & Chomsky's writings– or patriarchal institutions – Federici's, Butler's and Woolf's texts–. We targeted these authors because some of their books made part of our personal library in Spanish. The decision to extend our corpus to biographical and dystopian literature was based on the assumption that these types of texts would frequently consider 'what could have been but was not'. This assumption turned to be particularly true for Mandela's autobiography and Obama's biography on his father. Young activist Malala's and poet Audre Lorde's autobiographies enriched our corpus altogether with Bradbury's and Orwell's novels. An overview of the original titles and the translations analyzed is given in the Annexes.

---

[3] Other periphrases expressing some degree of probability in Spanish are *Haber que* + Infinitive and *Poder* + Infinitive (e.g. *Habría que haber traído uno aquí*, Should we have brought one of those here, [RAE 2009: 2148]; *Eva Girón salió de aquella viudedad con una alegría que, en todo caso, pudo haber sorprendido a quienes no conocían bien el estado de su relación con Umbrosa*, Eva Giron escaped from her widow condition with a joy that could somehow have surprised those who did not know the state of her relationship with Umbrosa [RAE 2009: 2153]).

[4] The Spanish reference corpus CREA contains 154 million words coming from 140.000 written documents published between 1975 and 2004. About 50% of the CREA's documents come from different types of manuscripts and the other 50% from press articles. Different varieties of Spanish are represented, with 50% of documents coming from diverse American varieties (Antilles, Chili, Mexico, Argentina, USA, among others), and 50% from Spain.

The freeware toolkit AntConc was used to identify the contextualized occurrences of *Should have* + Past participle. We elicited 95 sentences containing this item. We stored every single sentence carrying the targeted item within a table in a Word file. For the sake of transparency, every occurrence was given a code composed by the initials of the author plus a figure (e.g., VW#01, BO#95). This code will be useful in the Results section in order to quote or comment some of our examples. Working with AntConc allowed us to quantify the occurrences of *Should have* + Past participle per manuscript and thus, to have a picture of the authors who used it more and less frequently. Variation among authors will be further discussed in the Results section.

In general, we proceed to identify the equivalent translations in Spanish by using AntConc although a small part was done manually[5]. Every translated sentence was stored in a segmented way, right after the English source sentence. Then we analyzed the grammatical features of the Spanish segment corresponding to *Should have* + Past participle (e.g., tenses and verbal moods of the 95 translations, presence or absence of deontic markers, word order) and dumped the data to an Excel file. Finally, the Spanish translations were crossed with the online machine translations provided by Google[6].

## 4. Results

Translation segments generated by machines resulted more homogeneous in the syntactic constructions provided. Conversely, human output translation contained larger diversity of syntactic solutions, as illustrated in Figure 1. Human-generated translations presented heterogeneity of tenses covering the conditional, other indicative tenses, as well as subjunctive tenses. Unlikely, the segments generated by online machine translations were significantly rooted in the word-by-word formal equivalent in Spanish (i.e. the conditional *Debería haber* + past participle)[7]. Within the human translated segments, no significant differences were revealed by Chi-squared analyses. This means that the distribution of the patterns yielding the output constructions was not polarized but well balanced, as shown in Table 1[8]. For the sake of clarity, examples corresponding to the patterns summarized in Table 1 are provided below (10), (11), (12). Chi-square analyses revealed that the variation of means found in human outputs was significantly higher, compared to the machine outputs $(X^2 (4, N = 95) = 81.361, p < 0.001)$.

(10) Todo debería haber sido diferente en Rosario. [Cond HPP / Vb+H+PP]
(11) La promesa debía haber bastado. [IndHPP / Vb+H+PP]
(12) —¡Hubieses tenido que verla, Millie! [SubjPP / Aux+PP]

Table 1 shows four syntactic means in which the perfective *Should have* was translated into Spanish. The word-for-word translation *Debería haber* + past participle (in the graph, CondHPP) was the most extended construction in machine output segments, with 87 occurrences. A related construction containing also the auxiliary *Haber* followed by a past participle (IndHPP) was only found in human translations, with 19 occurrences. The frequency of the pluperfect subjunctive constructions (SubjPP) was also dissimilar across the groups, with one occurrence found in machine translations and 21 in humans'. Table 2[9] shows a different display of the data, accordingly to the syntactic structure of the translated constructions with no mention to verbal tenses and no mood distinctions. From a compositional viewpoint, the constructions provided by machines and humans appeared to hold significant differences, $X^2 (3, N = 95) = 42.543, p < 0.001$. The most used structure by both groups, humans and machines, was the three-word construction composed by an inflected form followed by the infinitive auxiliary *Haber* plus a past participle (Aux+H+PP). Under this structure was accounted example (10), which can be considered as the Spanish formal equivalent of the perfective *Should have*, and example (11), which can be taken as an approximate formal equivalent. The structure exemplified in (12) –i.e. the functional Spanish equivalent pluperfect subjunctive– was accounted under the two-word structure composed by an inflected form of the auxiliary *Haber* followed by a past participle (Aux+PP). This type of structure was used by human translators less frequently than the former structure (Aux+H+PP) but still more frequently than other solutions. Figures 1 & 2 are graphical displays of the corresponding Tables 1 & 2.

---

[5] This was the case for Barack Obama's *Los sueños de mi padre* and Audre Lorde's *Zami: una nueva forma de escribir mi nombre.*

[6] The segments' translations were extracted on 25th October 2018.

[7] Within the machine's output segments, Pearson's Chi-squared tests with Yates' continuity correction unveiled significant differences between the pattern CondHPP and each one of the other patterns ($p < 2.2e-16$).

[8] Legend: Cond = conditional tense, Ind = indicative tense, Subj = subjunctive tense, H = infinitive of the Spanish auxiliary *Haber*, PP = past participle.

[9] Legend: Aux = auxiliary verb, PP = past participle, Vb = inflected verb, H = infinitive of the Spanish auxiliary *Haber*.

|         | CondHPP | IndHPP | SubjPP | Others |
|---------|---------|--------|--------|--------|
| Machine | 87      | 0      | 1      | 7      |
| Human   | 27      | 19     | 21     | 28     |

**Table 1.** Patterns yielding the output translated segments

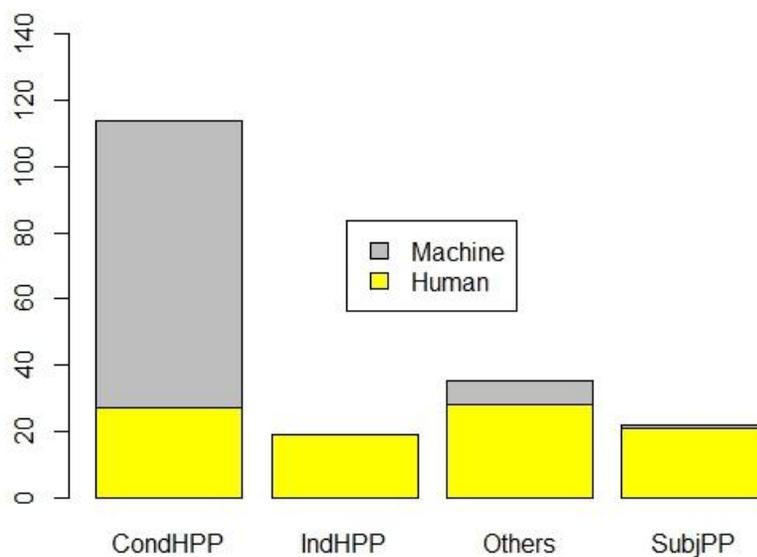|         | Aux+PP | Vb+H+PP | Others |
|---------|--------|---------|--------|
| Machine | 3      | 87      | 5      |
| Human   | 27     | 46      | 22     |

**Table 2.** Syntactic structure of the translated constructions

|         | Presence | Absence | p value[10] |
|---------|----------|---------|-------------|
| Machine | 91       | 4       |             |
| Human   | 73       | 22      | 0.0003      |

**Table 3.** Distribution of deontic markers across groups

|            | Indicative | Subjunctive | p value[11] |
|------------|------------|-------------|-------------|
| Machine    | 92         | 3           |             |
| Human[12]  | 68         | 26          | 7.78e-06    |

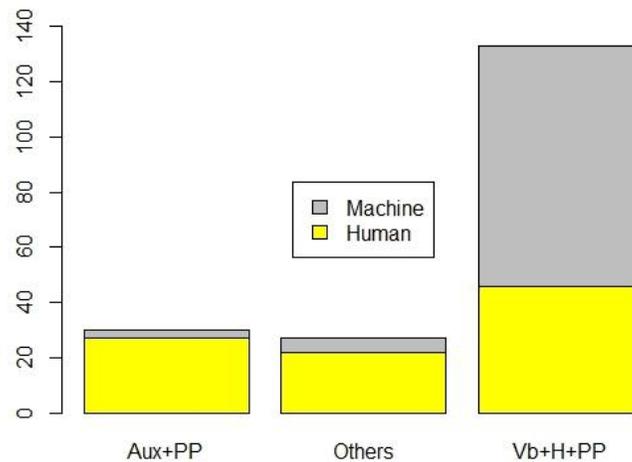**Table 4.** Distribution of mood across groups



**Figure 1.** Distribution of patterns across groups

---

[10] *p* value resulting from Pearson's Chi-squared test with Yates' continuity correction.
[11] *p* value resulting from Pearson's Chi-squared test with Yates' continuity correction.
[12] One occurrence excluded because non-inflectional form.

**Figure 2.** Distribution of syntactic structures across groups

Table 2 shows the distribution of deontic markers within the Spanish translations across groups. Less than 25% of the human translations were free-deontic constructions. In the case of machine translations this type of construction dropped under 5%. Notwithstanding, human translations preserved free-deontic constructions in a significant way, according to chi-squared analysis, $X^2$ (2, N = 95) = 12.878, $p < 0.001$). Similarly, human translations appeared to significantly preserve subjunctive constructions, compared to machine output translations, ($X^2$ (2, N = 95) = 19.991, $p < 0.001$). Frequencies of the latter constructions are presented in Table 4. Table 5 presents a general overview of the various structures found in the translations provided by machines and humans. Table 6 shows a summary of the repertoires of tenses and moods found across the groups. The conditional was overused in a significant way in machine translation outputs, whereas human translators significantly preserved the use of other indicative tenses and the subjunctive mood in their Spanish output segments (cf. chi-square analyses in Table 6).

| | Human | Machine |
|---|---|---|
| *Debería haber* + Past participle | 28 | 85 |
| *Debería / tendría que* + Infinitive | 5 | 2 |
| *Debería* | 0 | 2 |
| *Habrían* + Past participle | 2 | 0 |
| *Debí / debimos / debió haber* + Past participle | 5 | 0 |
| *Debí / debió / debieron* + Infinitive | 6 | 1 |
| *Debía haber* + Past participle | 13 | 0 |
| *Debía / debía de* + Infinitive | 3 | 0 |
| *Iba a* + Infinitive | 1 | 0 |
| *Podía* | 1 | 0 |
| *Han / hemos debido* + Infinitive | 2 | 0 |
| Bare present indicative (e.g. *regresa*) | 0 | 1 |
| *Hubiera/se debido* + Infinitive | 9 | 0 |
| *Hubiera/se* + Past participle | 12 | 1 |
| *Haya / hayan* + Past participle | 2 | 2 |
| Bare present subjunctive (e.g., *volviera*) | 3 | 0 |
| Others<br>– Adverbial constructions (e.g., *como debía, de lo que debía, en lugar de ir a parar*) | 3 | 1 |

**Table 5.** Classification of the translations across groups

| | Human | Machine | p values[13] |
|---|---|---|---|
| Conditional | 35 | 89 | 6.72e-16 |
| Other indicative tenses | 31 | 2 | 8.23e-08 |
| Subjunctive | 26 | 3 | 9.08e-06 |
| Adverbial constructions | 3 | 1 | 0.61 |

**Table 6.** Summary of the main tenses and moods across groups

## 5. Discussion

The linguistic variation of the constructions conveying *irrealis* scenarios in Spanish seem to be based in the combination of lexical and morphological features, as for example the presence or absence of modal verbs, and the use of verbal tenses from the indicative or the subjunctive. The constructions' word-order –i.e. three-word constructions *versus* two-word constructions– appears to be a dependent variable of the former features, suggesting that the use of the modal *Deber* (Should) introduces in Spanish a complex construction where the auxiliary' infinitive *Haber* plus a main's verb past participle seems to be mandatory, whereas the use of the free-modal pluperfect subjunctive implies simpler constructions. The human translation outputs are coherent with two main constructions well described in the Spanish grammar, such as the periphrasis *Deber* + Infinitive and the pluperfect subjunctive (cf. examples (7) & (2), respectively). However, a crucial difference between the grammar description and the output translations is the tense selected within the periphrasis construction *Deber* + Infinitive. On the one hand, examples provided by the RAE's grammar favors indicative tenses, as seen in (7) and (8), where *Deber* appears conjugated in the Spanish impefecto and the simple past, respectively. These constructions represented 25% of the translated output constructions provided by human translators, and was significantly underrepresented in the translations performed by machines, accounting for one occurrence. The most used tense within *Deber* + infinitive constructions were for both groups the conditional, which marked 29% output segments provided by humans and 89% provided by machines. Machines proved to significantly overuse the co-occurrence *Debería* + infinitive compared to humans, making of this construction a prominent way to translate the perfective *Should have*, no matter what temporal adverbs or aspectual instructions the English source segments carry. The consequences of the linguistic variation's impoverishment within the machine translated outputs are far to be calibrated and will depend on the use that professional translators and broader users do actually make regarding the MT's tools. One might expect the impact of machines' translations to be different in the case their outputs are just consulted, rather than actually read or assimilated.

The human intervention in post-editing MT outputs seem crucial here to avoid covert translations[14] inherited from the English morphosyntactic layer. In this sense, the mapping phase seems to play a crucial role during the translation task, and students should be trained in metalinguistic categories and crosslinguistic differences in order to prevent the anglizing impact of English in the target languages. In the mapping phase, contrastive knowledge is applied to all major linguistic, stylistic and cultural levels in order to map a source language form and content into that of target language (Kumar Jha 2004). This mapping is likely to be performed in a different, simplistic way by machines, at least for the perfective *Should have* in the English-Spanish configuration. In the machine translation outputs, the proliferation of the formal equivalent –i.e. *Debería haber* + past participle– and the invisibility of the functional equivalent –i.e. the free-modal pluperfect subjunctive– hence a reversal in the frequencies of the Spanish constructions pertaining to the counterfactual semantic domain. This might have a negative effect in the readability of the output segments by Spanish users, who may perceive their own language as invariably flattened.

## Conclusion

In the present survey evidence has been given about the cost of online machine translations in terms of linguistic variation in a corpus of Spanish counterfactual predicates. The frequency of the word-for-word translation *Debería haber* resulted to be significantly higher in the segments provided by machines, compared to humans. Conversely,

---

[13] *p* values resulting from Pearson's Chi-squared test with Yates' continuity correction.

[14] By the terms 'covert translation' or 'covert transfer' we intend the resulting construction from a subtle influence which leads to a form that is possible in the source language but does not correspond to what the native speakers normally produce in such a context. Covert transfers are hardly recognizable to the reader because the output construction is still grammatical. Nonetheless, covert transfers hence a shift at the level of frequency ratios between more or less frequent constructions pertaining to a specific semantic domain.

human translators preserved in a significant way a functional and more idiomatic equivalent construction in Spanish – i.e. the free-modal pluperfect subjunctive–. Our results suggest that the anglizing impact of English as a *lingua franca* (Palacios & Sánchez 2017) is significantly amplified when it comes to machine translation outputs. We hope to have drawn attention on the variation impoverishment that natural languages might incur if machine translation outputs are read or assimilated excluding any human post-edition mapping.

## References

Al-khresheh, M. H., Almaaytah, S. A. (2018). English Proverbs into Arabic through Machine Translation. *International Journal of Applied Linguistics & English Literature* 7(5): 158-166.

Conrad, S., Biber, D. (2001). Introduction: Multi-dimensional analyses and the study of register variation. In Susan Conrad & Douglas Biber (Eds.), *Variation in English: Multi-Dimensional Studies*, 3-12. Essex: Pearson.

Glowka, A. W., Lance, D. M. (1993). Preface. In Arthur Wayne Glowka & Donald M. Lance (Eds.), *Language Variation in North American English*. New York: The Modern Language Association of America.

Grevisse, M., Goosse, A. (2008). *Le bon usage. Grammaire française*. 13ème édition. Bruxelles: De Boeck – Duculot.

Kumar Jha, S. (2004). Human Vs. machine translatability of English-Hindi translation. *Bulletin of the Deccan College Research Institute* 64/65, 433-438.

Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry* 31, 231-270.

Palacios Martínez, I. M., Sánchez Fajardo, J. A. (2017). Introduction: Gaining insights into English as a contact language and its diffusion. *Alicante Journal of English Studies* 30, 5-22.

Lotz, S., Van Rensburg, A. (2016). Omission and other sins: Tracking the quality of online machine translation output over four years. *Stellenbosch Papers in Linguistics* 46, 77-97. doi: 10.5774/46-0-223

RAE (Real Academia Española) & Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española. Morfología Sintaxis I. Sintaxis II*. Madrid : Espasa.

Repiso, I. (2018). The impact of the source language in Spanish translations: A survey on English counterfactuals 'Should have'. *Meta: Translators' Journal* 63 (1), 139-159.

Repiso, I. (2014). La production des scenarios contrefactuels par des apprenants adultes hispanophones : quelques effets d'étrangeté liés à l'emploi du conditionnel en français langue étrangère. *Language, Interaction, Acquisition* 5(2): 249-278.

Schulz, S., Bernhardt-Melischnig, J., Kreuzthaler, M., Daumke, P., Boeker, M. (2013). Machine vs. Human Translation of SNOMED CT Terms. *Studies in Health Technology & Informatics*, 581-584. IOS Press. doi: 10.3233/978-1-61499-289-9-581

Siemund, P. (2011). *Linguistic Universals and Language Variation*. Berlin/New York: De Gruyter.

Verstraete, J. C. (2005). The semantics and pragmatics of composite mood marking: The non-Pama-Nyungan languages of northern Australia, *Linguistic Typology* 9, 223-268.

Van linden, A., Verstratete, J. C. (2008). The nature and origins of counter factuality in simple clauses. Cross-linguistic evidence. *Journal of Pragmatics* 40, 1865-1895.

Vetters, C. (1994). A propos de l'imparfait après si-. In: Marc Plenat, Michel Aurnague & Anne Condamines (Eds.), *L'emprise du sens. Structures linguistiques et interprétations*. Amsterdam: Rodopi, 337-355.

Toral, A., Wieling, M., Way, A. (2018). Post-editing Effort of a Novel With Statistical and Neural Machine Translation. *Frontiers in Digital Humanities* 5:9. doi: 10.3389/fdigh.2018.00009

Yamazaki, S., Sigley, R. (2013). Preface: Approaching Variation. In Shunji Yamazaki & Robert Sigley (Eds.), *Approaching Language Variation though Corpora*, 17-21. Bern: Peter Lang.

**Annexes:** Corpus' source texts and translations

| | Original | Translation |
|---|---|---|
| AL | Lorde, A. (1982). *Zami: A New Spelling of My Name*. London: Persephone Press. | *Zami: una nueva forma de escribir mi nombre*. (2010). Traducción de María Durante. Madrid: Horas y horas. |
| BO | Obama, B. (2004). *Dreams from My Father*. New York: Three Rivers Press. | *Los sueños de mi padre*. (2008). Traducción de Fernando Miranda y Evaristo Páez Rasmussen. Almed: Granada. |
| DK | Kahneman, D. (2011). *Thinking fast, thinking slow*. London: Penguin. | *Pensar rápido, pensar despacio*. (2012). Traducción de Joaquín Chamorro Mielke. Debate. |
| GO | Orwell, G. (1949). *1984*. | *1984*. (1980). Traducción de Rafael Vázquez Zamora. Barcelona: Salvat. |
| JB | Butler, J. (1999). *Gender Trouble*. New York: Routledge. | *El género en disputa*. (2007). Traducción de María Antonia Muñoz. Barcelona: Paidós. |
| MY | Yousafzai, M., Lamb, C. (2013). *I Am Malala*. London: Orion. | *Yo soy Malala* (2013). Traducción de Julia Fernández. Madrid: Alianza. |
| NC | Chomsky, N. (2003). *Hegemony or Survival: America's Quest for Global Dominance*. New York: Metropolitan. | *Hegemonía o supervivencia: El dominio mundial de EEUU* (2004). Traducción de Carlos José Restrepo. Bogotá: Norma. |
| $NC_2$ | Chomsky, N. (1989). *Necessary Illusions. Thought Control in Democratic Societies*. New York: South End Press. | *Ilusiones necesarias: control del pensamiento en las sociedades democráticas*. (1991). Traducción de Loreto Bravo de Urquia y Juan José Saavedra Esteban. |
| NK | Klein, N. (2000). *No Logo*. Great Britain: Flamingo. | *No Logo*. (2001) Traducción de Alejandro Jockl. Barcelona: Paidós. |
| $NK_2$ | Klein, N. (2007). *The Shock Doctrine*. New York: Metropolitan. | *La doctrina del shock*. (2007). Traducción de Isabel Fuentes García, Albino Santos Mosquera y Remedios Diéguez Diéguez. Barcelona: Paidós. |
| NM | Mandela, N. (1994). *Long Walk to Freedom*. Boston: BackBay. | *El largo camino hacia la libertad*. (2012). Traducción de Antonio Resines y Herminia Bevia. Madrid: Aguilar. |
| RB | Bradbury, R. (1953). *Fahrenheit 451*. New York: Ballantine Books. | *Fahrenheit 451*. (2006). Traducción de Alfredo Crespo. Almería: Ediciones Perdidas. |
| SF | Federici, S. (2004). *Caliban and the Witch*. Autonomedia. | *Calibán y la bruja*. (2010). Traducción de Verónica Hendel y Leopoldo Sebastián Touza. Madrid: Traficantes de sueños. |
| VW | Woolf, V. (1929) A room of one's own. London: The Hogarth Press Ldt. | *Una habitación propia*. (2008). Traducción de Laura Pujol. Barcelona: Seix Barral. |