What got lost in online machine translations? Effects on Aspect and Passivization from a literary corpus

Isabel Repiso

What got lost in online machine translations? Effects on Aspect and Passivization from a literary corpus

Isabel Repiso

Salzburg University

Abstract: Aspect is a crucial ingredient of temporality, which reflects the speaker's internal perspective on a given situation (Ayoun 2013). This article compares human translations and online machine translations (MT) with the purpose of describing how the distinction between perfective and imperfective aspect is realized in Spanish when translating from English. Our results come from 1.6 million-word Social Sciences corpus from where 82 perfective *get*-passive constructions (e.g., *I got saved*) were elicited. Although the general pattern is for atelic predicates (states and activities) to occur with the imperfect tense and for telic predicates (accomplishments and achievements) to occur with the preterite, in Spanish all the aspectual predicates can be expressed with preterite and imperfect, depending on what the speaker wants to convey (Montrul & Slabakova 2003: 357). The results show that Spanish imperfect readings were largely preserved in human translations compared to online machine translations. This study discusses the machines' overextension of the preterite as a systemic regularity that bring MT's outputs close to L2 learners' grammars.

Key-words: Grammatical aspect, *get* passive, machine translation, Spanish.

Introduction

Aspect is a language property that has been at the center of relevant cross-linguistic studies in the domain of language acquisition (Benazzo & Andorno 2017, Dosi et al. 2017, Ayoun 2013,

Mayberry 2011, Cuza 2010, Montrul & Slabakova 2003). In general, these studies have aimed to describe how L2 learners and heritage speakers perform grammatical and lexical Aspect[1]. This question has been crucial to gaining insight in the complex process of Form-to-function mapping. Beginner Spanish learners have given evidence on the salient use of achievement verbs in perfective tenses and state verbs in imperfective tenses (López-Ortega 2000). To date, the realization of grammatical Aspect by machine translation remains a relatively unexplored question.

The goal of this study is to compare the output translations made by the online Google machine and by human translators at the syntactic level for a set of undergoer arguments[2] carrying the *get*-passive (e.g., *he got scared*). The analysis will account for the grammatical aspect or the internal temporal constituency of a situation (Comrie 1976) in the Spanish outputs. In contexts relating to the past, perfective and imperfective readings are available in English by one morphological form (*-ed*), whereas Spanish encodes this contrast by selecting a specific past tense (e.g. the perfective *cantó* or the imperfect *cantaba*). This study will explore the translation of *get* passive constructions related to the past (i.e. *got* + past participle constructions). In English this construction usually entails a semantic value pertaining to accomplishments or achievements (e.g, *He got shot; They got rescued*). Because of this semantic adscription, its translation may be particularly problematic in Spanish, where the equivalent construction is not limited to a certain

[1] Grammatical Aspect is expressed through morphological markers. French conveys perfective readings by means of the *passé compose* and the *passé simple*, and imperfective readings by means of the *imparfait* and the *plus-que-parfait*. Spanish conveys perfective readings by means of the *pretérito indefinido (Juan leyó)*, and imperfective readings by means of the *pretérito imperfecto (Juan leía)*. Lexical Aspect depends on the semantic properties [±dynamic, ±telic, ±punctual] expressed by the utterance's type of verb (state, activity, accomplishment, achievement).

[2] From a syntactic viewpoint undergoer arguments refer to the participants who are viewed as primarily affected in a given state of affairs (Van Valin & LaPolla 1997: 149).

type of aspectual instruction but instead allows a wider range of viewpoints from which the action can be presented. In addition, our analysis will describe the syntactic means by which the passive voice is realized in the output translations. We assume the nature of English as a productive language for conceptualizations in the passive voice, compared to Spanish. Precedent studies have shown that the realization of passivized information is expressed in English by periphrastic passive constructions containing either the auxiliaries *be* or *get*, and in Spanish by the reflexive passive which is characterized by the lack of a specified agent, an active form of the verb, and the reflexive morpheme *se* (Agnieszka 2012)[3].

Our research questions are as follows:

a. How do machines handle imperfective readings in their Spanish outputs as compared to humans? Because English lacks a past tense analogous to the imperfect it may ~~well~~ be that online machine translations will overextend the perfective in Spanish to contexts where the imperfect is preferred;

b. Are human translators more sensitive to imperfective readings, and thus more likely to preserve them in the translated Spanish outputs?

This article is organized in six sections. The first section presents an overview of preceding studies that compared machine to human translations. In the subsequent two sections, the cross-linguistic differences between Spanish and English are explained regarding the perfective-imperfect contrast (section 2), and the use of the passive voice (section 3). Section 4 introduces the literary works or corpus from which the *get*- passive constructions were taken. In section 5,

[3] For example: *La semana pasada se subió el precio del pan* (Last week the price of bread went up).

we examine our results in light of the passive/active voice and perfective/imperfective contrast. Section 6 ties our results to preceding studies dealing with online machine translations.

1. Preceding studies dealing with machine translations and Aspect

Translation universals are transversal phenomena playing a role in the production of hybrid texts that partially correspond to the source text and partially to texts originally written in the target language. In this sense, every translated text is the product of a certain degree of interference and a certain degree of standardization on various levels. Interference pertains to the features of the source text that are readily recognizable in the translation output. Standardization pertains to the textual relations from the original that are modified to the point of being ignored, in favor of habitual options offered by a target repertoire (Toury 1995). Preceding studies that compared human and machine translations are based on sophisticated quantitative analyses covering a wide range of text features such as sentence length, punctuation or type/token ratio.

At the syntactic level, Volansky et al. (2011) provide evidence of the similarity between machine translations (MT from the forward) and original texts for the ratio of passive forms for all verbs. The passive verb ratio was slightly dissimilar when comparing human and machine translation, thus indicating the specificity of human translators concerning standardization procedures accordingly to the target language's syntactic patterns[4]. The comparison between Google's machine translations and human translations has also proved that the latter are closer to each other than to Google (El-Haj et al. 2014). These authors confirmed that some language configurations prompted higher accuracy ratios when translated by Google's machine. This was

[4] Two corpora were used by Volansky et al. (2011) in their survey. The first contained (a) original English journalistic texts, and (b) English outputs translated from Hebrew journalistic texts. The second was a machine translation corpus consisting of English outputs translated from Hebrew journalistic texts. As for this specific language configuration, Volansky et al. (2011) assumed that the passive voice is more frequent in English.

the case for French source language/English target language, compared to Arabic target language. The goal of comparing the distance between the human translations and MT output is to test the assumption of reference, which states that the closer the machine translation is to a professional human translation, the better it is (Papineni et al. 2002). However, it has been claimed that distance-based evaluation methods are somehow problematic because mismatches do not necessarily mean degradation in MT quality (Babych and Hartley 2004).

As mentioned in the Introduction, Aspect is a relatively unexplored language property in the field of Translation studies but very well defined from a linguistic viewpoint. The notion of Aspect is at the center of the Discourse Representation Theory (Kamp 1981) –which accounts for the explanation of narrative texts–, and the Failed Functional Features Hypothesis (Hawkins and Chan 1997) which predicts that the L2 learners' grammars will deviate from those of native speakers. Both notions will be approached as explanatory factors of our results in the Discussion. In the Discourse Representation Theory the meaning of a sentence contributes to the meaning of the text or discourse. According to this Theory, the aspectual meaning of a sentence is a composite of the information from the components of viewpoint and situation type (Smith 2012). The parameter of viewpoint deals with progressive, perfective and imperfect interpretations. Our results should test whether the lack of perfective/imperfect contrast involving past tenses in English make Spanish outputs less idiomatic by favoring perfective readings and limiting imperfect ones. The interest of discussing the Failed Functional Features Hypothesis in light of our results is to consider the typological differences between English and Spanish as an explanation for the differences between machine and human translators. In the same way learners are not able to incorporate certain functional features to their L2 grammars, the eventual overextension of perfective readings in the Spanish outputs would likely produce deviate texts.

2. The *get*- passive in English and its equivalent construction in Spanish

In English, the *get* passive has traditionally been compared to the canonical *be* passive in order to explain and distinguish its respective uses. The present section highlights the preferred contexts where the construction *got* + past participle is used in English and its semantic restrictions (point 2.1), as well as the equivalent constructions in Spanish (point 2.2). Cross-linguistic differences will be discussed in order to identify the morphosyntactic features that may be problematic in translations (point 2.3).

2.1 Empirical remarks about the *get*- passive in written English

The *get*- passive has been portrayed as highly register-sensitive, as a marker of spoken rather than written, and colloquial rather than formal genres (Anderwald 2018). Recent research has shown clear differences in the way *get*- and *be*- passives are used in American English (Schwarz 2018). More specifically, a diachronic survey of the TIME Magazine Corpus shows that the use of *get*- passives has progressively increased within accomplishment predicates over the last century (Schwarz 2017: 321). However, the highest frequencies of *get*- passives are still associated with transitional act predicates, as has been the case from 1920. The following examples, from Schwarz (2017: 318), illustrate how perfective *get*- passives are used in accomplishments predicates and in transitional acts predicates, respectively.

(1) The grid city of 19th century Barcelona, designed from the ground up as an ideal townscape by the socialist engineer Ildefons Cerda, is the biggest example of would-be Utopian town design that ever **got built** – but neither it nor its inventor rates a mention in the catalog. (TIME 2000s)

(2) I never **got arrested** like Robert Downey Jr. – more because of dumb luck or chance than anything else. (TIME 2000s)

In general, *get*- passives illustrate the result of an event or state-of-affairs which directly affects the subject and can be interpreted from a pragmatic point of view as a reflexive construction (Agnieszka 2012: 179). Sometimes using the *get*- passive enables the author to focus on the event as such and to convey the reflexive character of the sentence so that the subject is presented as agent and patient, like in example (3). This construction can also serve to emphasize the result of the process, as well as the subject's influence or will to perform the action (4). In examples 3 and 4, from Agnieszka (2012: 180), the auxiliary *get* can be replaced by *be*, and in that case no influence on the part of the subject is assumed. The Spanish equivalents to the construction *got* + past participle would be reflexive active constructions or impersonal active constructions. In the examples that follow, the bold segments may be translated as *me mojé* and *lo enviaron*, respectively (Agnieszka 2012: 179).

(3) I ran a bit and walked a bit, but it was still pouring with rain and I **got soaked** (*The Sun*, "Lorraine Kelly is the face of *The Sun's* Fit Squad Diet", 08/04/2010)

(4) Maybe he **got sent** to Nigeria (*Guardian*, "Blood and oil and panorama: passports to kill", 30/03/2010)

The examples presented so far topicalize the patient as an undergoer entity, that is, the participant the speaker presents is affected most by the action (Van Valin & LaPolla 1997: 145). Therefore, the perfective *get*- passive can be considered ~~as~~ a productive construction within contexts expressing some degree of affectedness[5]. As for speech data, Thompson et al. (2013) claim that the selection of *get*- passives over *be*- passives works as a focus marker used by the speaker to highlight the patient's role.

2.2 Equivalent constructions in Spanish

Affectedness in Spanish may be expressed in diverse ways but we will focus on three syntactic means: active-voice sentences in which the agent is not specified and that may be introduced by an accusative complement (example 5); canonical *ser* passive sentences (6); and causative sentences containing the reflexive *hacerse* + infinitive (7).

(5) A Juan lo agredieron

To Juan him they attacked

(6) Juan fue agredido

Juan was attacked

(7) Juan se hizo maquillar

Juan made himself makeup

2.2.1 Active-voice sentences

[5] In this paper affectedness is intended as a semantic parameter encoding a persistent change in – or impingement of – an event participant.

In example (5), the brackets indicate an informational segment that is implicitly understood by the hearer by means of the clitic pronoun *lo*, which marks the accusative. The particularity of Spanish in marking a case by means of the morpheme *a* can be explained diachronically as a feature inherited from Latin (Gauchola 2012). In Spanish, the marker *a* is mandatory to introduce accusatives involving animated entities (RAE 2009: 2226). The co-occurrences of the clitic pronoun *lo* plus the accusative introduced by *a* might be considered redundant within the same sentence.

2.2.2 *Ser*- passive sentences

Example (6) is representative of canonical *ser*- passive sentences in Spanish, in which the past participle coordinates in number and gender to the subject. In this kind of construction the subject nominates the verb action's patient (RAE 2009: 3041). Although the canonical *ser*- passives and the active-voice sentences can convey similar informational contents in Spanish, the selection of one over another seems to be context dependent. Preferring a periphrastic passive construction may be due to the speaker's will to emphasize the patient role and relegate the agent's to a background level of information, or it may respond to a lack of information concerning the agent (RAE 2009: 3042).

2.2.3 Causative sentences containing *hacerse* + infinitive

An important formal element of this construction is the morpheme *se*, which in general works as a pronoun allowing the verb's action to be read as reflexive and thus, linked to the nominal predicate. Interestingly, this is not the case in example (7), where the idiomatic interpretation is that someone else made the subject's makeup (Bosque & Demonte 1999). A way of paraphrasing

example (7) is example (8). Both (7) and (8) have identical meanings to Spanish listeners. Like in English, this type of construction allows intransitive verbal predicates containing infinitive forms followed by an accusative, as in example (9). In passive readings, infinitive forms may be followed by agent complements introduced by prepositions *por* (by) or *de* (of)[6].

(8) [Él] Hizo que lo maquillaran

He made sure to be made up

(9) Se hizo servir el café en la biblioteca (RAE 2009: 1990)

He got some coffee served at the library

As seen in examples (7) and (9), *hacerse* + infinitive applies to active readings in Spanish where the subject's construction corresponds to the causative action's agent (Glushchuk-Oleia, 2014: 148). Patient readings where the subject is presented as the undergoer of non-volitional actions are possible by means of the construction *dejar* + infinitive[7]. Interestingly, the general description according to which *hacerse* + infinitive is semantically restricted to active readings does not apply to some Spanish-speaking communities in contact with Amerindian languages like Quechua or Aymara. In some rural populations of Ecuador it has been found that the construction *hacer (de)* + infinitive can be combined with intransitive verbs to express patient readings[8].

[6] Examples: *Se hizo perdonar por el gobernador* (He was forgiven by the governor) or *Se hacía acompañar de una orquestina* (He was accompanied by a small orchestra) (RAE 2009: 1990).
[7] Examples: Se dejó morir, *[He/She] let himself/herself died*; No te dejes caer, *Do not let yourself fall* (RAE 2009: 1991).
[8] Example: Me hice de caer = Me caí; *I got myself fallen = I fell* (RAE 2009: 2014).

2.3 Cross-linguistic differences

The comparison between the realization of English *get*- passives and Spanish equivalent constructions yields some important points in the conceptualization of affectedness. From a formal perspective, *got* + past participle has a syntactic cognate in Spanish, which is the construction *hacerse* + infinitive. Importantly, the Spanish *hacerse* does not cover the same semantic values as its English counterpart, since it can only topicalize patients as benefactors but not as undesired actions' undergoers (e.g. *Se hizo disparar, *He got shot*). This semantic restriction of *hacerse* makes its use contextually limited compared to the canonical *ser*- passives and the active-voice constructions, which in turn can topicalize patients within non-volitional predicates. The topicalization of a different participant from the agent can be conveyed in Spanish by others means than the passive voice using a bare active construction (e.g. Lo agredieron, *He was attacked*).

3. Theoretical assumptions: the perfective/imperfect contrast

A crucial difference between Spanish and English is that the perfective/imperfective contrast is not grammaticalized in English. In Spanish perfective and imperfective meanings are conveyed in the grammaticalized inflectional morphology. In past-time contexts, the aspectual viewpoint of an action can be presented as bounded in time (10) or as unbounded (11)[9].

(10) María tocó el piano.

Maria played-PRET the piano

'Maria play*ed* the piano.'

[9] Examples from Cuza (2010).

(11) María toca*ba* el piano cuando era niña.

Maria played-IMP the piano when child

'Maria play*ed* the piano as a child.'

Unlike Spanish, in English both aspectual notions are available by means of the simple past tense (*-ed*). Cross-linguistic studies have shown that this morphosyntactic difference has critical implications for the interpretive domain (Mayberry 2011), and this is seen in Table 1, which summarizes some examples from our corpus (cf. Appendix).

Author[10]	Original	Human translation	Machine translation
MY	When we entered the shop we were going to, the shopkeeper laughed and told us he **got scared** thinking we might be suicide bombers as many suicide bombers wore the burqa.	Cuando entramos en la tienda a la que íbamos, el dueño se rio y nos dijo que **se había asustado** por si éramos terroristas suicidas, porque muchos terroristas suicidas se ponían un burka.	Cuando entramos en la tienda a la que íbamos, el tendero se echó a reír y nos dijo que **se asustó** al pensar que podríamos ser terroristas suicidas, ya que muchos terroristas suicidas llevaban el burka.
NK	She asked us to carry a message to everyone who was trying to help the tsunami survivors. "If you	Nos pidió que si podíamos llevar un mensaje a todos aquellos que estaban intentando ayudar a los	Nos pidió que lleváramos un mensaje a todos los que intentaban ayudar a los sobrevivientes del tsunami. "Si

[10] Full names of authors can be found in the Appendix. MY = Malala Yousafzai; NK = Naomi Klein; NM = Nelson Mandela.

	have something for me," she said, "put it in my hand." Sri Lanka wasn't the only country that **got hit** by this second tsunami — similar stories of land and law grabs have come out of Thailand, the Maldives, and Indonesia.	supervivientes del tsunami. "Si tienes algo para mí", dijo, "ponlo en mi mano". Sri Lanka no era el único país que **había sido golpeado** por este segundo tsunami. Historias similares de tierras y expropiaciones se revelaban en Tailandia, las Maldivas e Indonesia.	tienes algo para mí", dijo ella, "ponlo en mi mano". Sri Lanka no fue el único país que **se vio afectado** por este segundo tsunami. Historias similares de acaparamientos de tierras y leyes surgieron de Tailandia, Maldivas e Indonesia.
NM	Nevertheless, small problems sometimes **got blown** out of proportion. In 1983, during a visit with Winnie and Zindzi, I mentioned to my wife that I had been given shoes that were a size too small and were pinching my toe. Winnie was concerned, and I soon learned that there were press reports that I was having a toe amputated.	No obstante, las pequeñas dificultades a veces **se veían magnificadas**. En 1983, durante una visita de Winnie y Zindzi, le comenté a mi esposa que los zapatos que me habían dado me estaban pequeños y me apretaban. Winnie se mostró preocupada y no tardé en averiguar que la prensa había publicado la noticia de que iban a amputarme un dedo del pie.	Sin embargo, los pequeños problemas a veces **se desbordaron**. En 1983, durante una visita con Winnie y Zindzi, le mencioné a mi esposa que me habían dado zapatos que eran demasiado pequeños y me estaban pellizcando el dedo del pie. Winnie estaba preocupada, y pronto supe que había informes de prensa de que me estaban amputando un dedo del pie.

Table 1. Comparison between human and machine output translations

Table 1 highlights some differences between human and machine output translations regarding the grammatical aspect or viewpoint from which events are presented. From a discursive viewpoint, imperfective aspect marks background information significantly whereas foreground information is expressed significantly more often with perfective aspect marking (López-Ortega 2000). This is the case in the examples from Table 1 that were done by human translators. If we take a look at Malala Yousafzai's segment, we can see that the imperfect *se había asustado* (got scared) forms part of a subordinate relative clause which works as the direct object of the perfective *dijo* (he told). Naomi Klein's *got*- passive is preceded by a touching fragment where she introduces a tsunami survivor called Renuka. The narrative focus is put on Renuka's story and many specific details related to the past are provided, such as the death of her elder children and how she managed to escape the wave despite being nine months pregnant. In the Spanish translation, the preterites –i.e., *pidió*, *dijo* (she said, she asked)– are used to mark the text's main structure and the imperfective *había sido golpeado* comes subsequently as a side structure to enlarge the picture and defocus from Renuka's drama. The opposite pattern is found in Mandela's translation, where a general frame is first introduced by the imperfective *se veían magnificadas* (got blown out of proportion) to come to a specific anecdote happened in 1983 which, in turn, is marked by a reiterative use of the preterite (i.e., *comenté*, *mostró*, *tardé*). This narrative distinction between the main and side text' structures by means of the perfective/imperfective contrast usually modulates Spanish writing texts. If we take a look at the online translations generated by machines we can see that the narrative skeleton is somehow flat and poor in discursive edges (cf. Table 1, third column). This is essentially true in the translation of Klein's fragment where no contrast is available to the Spanish-speaking reader because all the verbal forms have been transformed in simple past occurrences.

4. Corpus and methods

The corpus used to run the study was based on 13 literary works written by 11 authors. Many of the titles dealt with identity and racial issues. This is the case for Toni Morrison's novels *Beloved* and *Sula*, Barack Obama's biography, and the autobiographies by Audre Lorde, Nelson Mandela and Malala Yousafzai. Two titles dealt with the effects of capitalism and globalization –Naomi Klein's *No Logo* and *The Shock Doctrine*–, and two titles depicted dystopian societies characterized by totalitarian regimes –George Orwell's *1984* and Ray Bradbury's *Fahrenheit 451*–. These 10 works can be considered representative of the Social Sciences domain, combining historical facts and critical views.[11] Our Social Sciences corpus was completed by 3 works written before 1930 whose digitalized versions were readily available both in English and Spanish – Emily Brontë's *Wuthering Heights,* Jane Austen's *Pride and prejudice,* and Scott Fitzgerald's *The Great Gatsby*. The presence of the source constructions 'got + past participle' was made automatically using the freeware toolkit AntConc. The presence of the target constructions was also made automatically for 11 translated titles, and manually for 3 translated titles from printed books[12]. Statistical analyses and graphs were run and generated using R from original Excel spreadsheets previously converted in delimited text files (.txt).

5. Results

[11] One of the main concerns when selecting the works for our corpus had to do with the frequency of the *get*- passive construction. In English, the *get*- passive has been described as a construction typically produced in speech by children up to 5 years old (Leonard et al. 2003). The idea of building our study based on 13 Social Sciences literary works came from the assumption that biographies and essays are subjective type of texts in which speech productions or dialogues are more likely to occur. In all, 69 *get*- passive constructions were chosen from this corpus. In the aim of running statistical analyses on a larger number of *get*- passive constructions, we decided to include three novels whose texts were available online. The addition of these works allowed us to obtain 82 *get*- passive constructions.

[12] Audre Lorde's *Zami*, Barack Obama's *Los sueños de mi padre*, and Toni Morrison's *Beloved*.

In all, 82 '*got* + past participle' constructions were analyzed. For the purpose of our survey, we only considered reflexive constructions, excluding those in which the past participle's subject does not match with the subject of the perfective *get*[13]. The comparison between the translations performed by humans and those performed by machines did not reveal significant differences concerning the sentence's voice[14]. Both groups provided active voice constructions prominently over passive voice sentences in Spanish, representing 73% in the human group and 84% in the machine group (cf. Table 3). The perfective/imperfective contrast proved to be significantly different when comparing the translations generated by humans and those generated by machines[15]. The former preserved imperfective readings in 32% of their translations, whereas the latter did it only in 14% (cf. Table 2). Figure 1 shows the distribution of perfective aspect *versus* imperfective aspect across the translations performed by humans and machines (left side), and the distribution of the passive-voice *versus* active-voice constructions (right side).

	Human translations	Machine translations
Perfective	47	67
Imperfective	27	12
Others	8	3

Table 2. Perfective-imperfective contrast across groups

	Human translations	Machine translations

[13] 'I've **got** my wife **lockedin** up there,' explained Wilson calmly. (*The Great Gatsby*, Scott Fitzgerald). The reason why these kinds of construction were not considered is that they are not equivalent to the Spanish passives where the past participle always refers to the subject. In the example above, "locked in" refers to the wife and thus it could be translated as a bare adjective (e.g., *Tuve a mi mujer encerrada ahí arriba*).

[14] *p*-value = 0.2 resulting from the Welch two sample t-test analysis.

[15] *p*-value = 0.002 resulting from the Welch two sample t-test analysis.

Active	60	69
Passive	15	10
Others	7	3

Table 3. Passive voice *versus* active constructions across groups

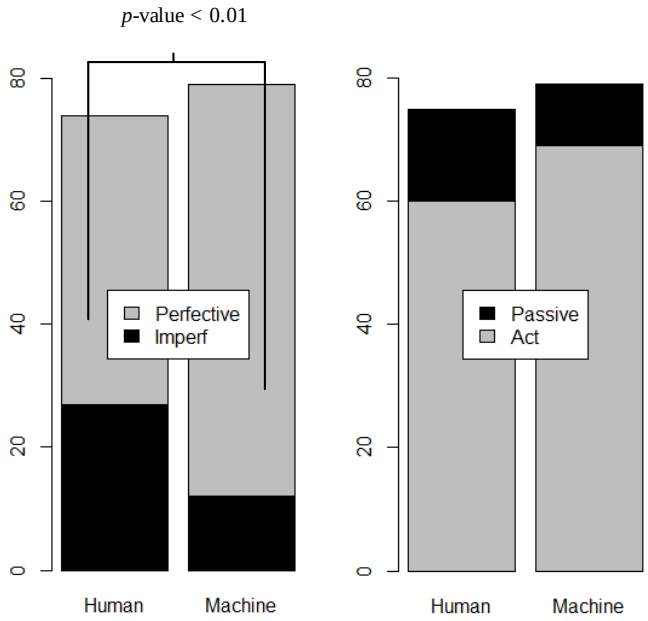

Figure 1. Spanish output constructions across groups

Some *got-* constructions were translated the same way by humans and machines when conjugated with certain verbs. This was the case for the collocates *got wind*[16] and *got involved*[17]. Surprisingly, the construction *got married* –four occurrences– was translated only twice using

[16] Three occurrences were elicited from the English corpus. These occurrences were translated either by humans or machines using the simple past in Spanish (i.e., *se enteró*, *se enteraron*).

[17] Four occurrences were elicited from the English corpus. Three of them were translated using the simple past in Spanish both by humans and machines. However, some variation was found at the translations' lexical level. The outputs provided by machines always contained the verb *involucrarse*, whereas the human translators opted for *dedicarse*, *incorporarse* and *participar*. The fourth occurrence of *got involved* was graciously omitted by an ellipsis in the human translation and translated again as a simple past in the machine translation.

the simple past both by humans and machines (i.e., *me casé, nos casamos*)[18]. Figure 2 presents the frequencies of the *get* passives per author[19] and gives an overview on the behavior of the two groups across the 82 source constructions analyzed. In particular, it illustrates the common and dissimilar translations provided by humans and machines[20]. The x axis presents the distribution of the 82 *get* passives across the 13 source texts of our corpus. The y axis represents the matching frequencies crossing the translations provided by each group (e.g., whenever both the human translator and the machine have produced a simple past, or whenever they have both produced an imperfective construction). Three main patterns were found when comparing the MT outputs to the human translators' outputs. When both the machine and the human translator provided a past simple in Spanish the *get* passive was ranged under the label Equal_SP (black bars, Figure 2). When both the machine and the human translator provided an imperfective past in Spanish the occurrence was ranged under the label Equal_IMP (pink bars, Figure 2). When the human translator provided an imperfective form and the machine provided a perfective form the occurrence was ranged as Imp_Vs_Perf (grey bars, Figure 2).The pattern Equal_SP corresponded to 50% of the total *get* passives, the pattern Equal_Imp corresponded to 3.6% and the pattern Imp_Vs_Perf corresponded to 31.7%.

[18] The other two occurrences of *got married* were dissimilarly translated by humans and machines. The latter generated simple past constructions, whereas the former provided imperfective constructions (i.e., *¿te casaste?* versus *¿estás casada?*, and *se casó* versus *se había casado*).

[19] AL = Audre Lorde, BO = Barack Obama, EB = Emily Brontë, GO = George Orwell, JA = Jane Austen, MY = Malala Yousafzai, NK = Naomi Klein, NM = Nelson Mandela, RB = Ray Bradbury, SF = Scott Fitzgerald, TM = Toni Morrison.

[20] Legend: Equal_IMP = both translations carried imperfective constructions, Equal_SP = both translations carried simple past forms, Imp_Vs_Perf = the human translation carried an imperfective construction and the machine translation carried a simple past form, Others = others cases.

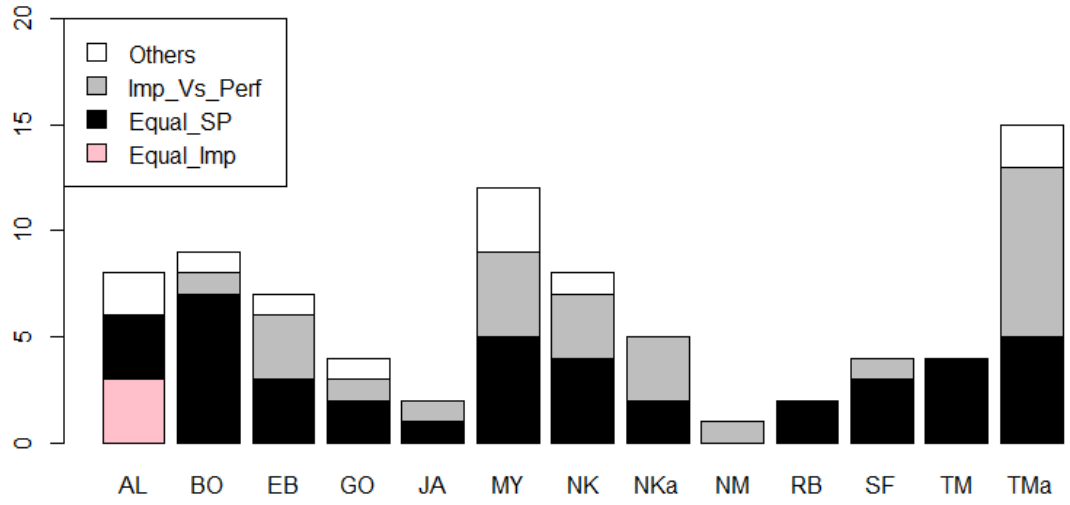

Figure 2. Patterns emerging after crossing human and machine translations

As shown in Figure 2, the majority of constructions used in the human translation of Obama's biography matched the online machine outputs, with a large amount of past simple forms. On the opposite side of the figure, the translators of Morrison's *Beloved*, Malala's autobiography, and Klein's *The Shock Doctrine* differ from machine outputs in their wider use of imperfective constructions. The case of Lorde's autobiography is striking, with some imperfective constructions generated by the translation machine that match with the human outputs (cf. Table 4, third example). The larger preservation of imperfective readings by human translators –cf. Figure 2's gray bars– hence the machines' failure to restitute complex discourse dynamics in the target language and, more particularly, to overcome the parametric features available in the source language.

Table 4 summarizes some examples of the main patterns captured in Figure 2. In Obama's fragment, the source construction belongs to a reported speech that captures the positive reaction of a school principal when he talks about the likelihood of perceiving funds for a pilot program. The translation in Spanish by a past simple construction seems perfectly congruent with the

discursive situation, in which a young Obama talks to Johnnie Owens about a precise moment related to the past. In the fragment from *Beloved*, the narrator tells what has been revealed to her about her inner nature. The human translation is subtle and finest compared to the machine output, not only syntactically but also lexically.

Author	Original	Human translation	Machine translation	Pattern
BO	"Oh, he was all smiles," Johnnie said. "Said he really liked the proposal. He **got** real **excited** when he heard we might get funding.	-¡Bueno! Se deshizo en sonrisas –dijo Johnnie-. Se mostró encantado con la propuesta. Y **se entusiasmó** cuando supo que podríamos conseguir fondos.	"Oh, él era todo sonrisas", dijo Johnnie. "Dijo que le gustó mucho la propuesta. **Se emocionó** mucho cuando escuchó que podríamos obtener financiamiento.	Equal_SP (n = 41) 50%
TM	That I was charmed. My birth was and **I got saved** all the time. And that I shouldn't be afraid of the ghost. It wouldn't harm me because I tasted its blood when Ma'am nursed me.	Dijo que yo estaba encantada. Que mi nacimiento había sido obra de un encanto y que siempre **me salvaba**. Que no debía tener miedo del fantasma. Que no me haría daño porque había probado su sangre cuando me dio la teta.	Que estaba encantado. Mi nacimiento fue y **me salvaron** todo el tiempo. Y que no debería ten miedo del fantasma No me haría daño porque probé su sangre cuando Ma'am me cuidó.	Imp_VS_Perf (n = 26) 31.7%
AL	Sometimes I dabbed	A veces le daba con él a	A veces, frotaba las	Equal_Imp

	the figures on either side of the head behind the ears as I had seen my mother do with her glycerine and rosewater when **she got dressed** to go out.	las figuritas unos leves toques a ambos lados de la cabeza, detrás de las orejas, como le había visto hacerlo a mi madre con la glicerina y el agua de rosas cuando **se arreglaba** para salir.	figuras a cada lado de la cabeza detrás de las orejas, como había visto a mi madre hacer con su glicerina y agua de rosas cuando **se vestía** para salir.	(n = 3) 3.6%

Table 4. Summary of the translation patterns performed by humans and machines

6. Discussion

A crucial research question of this paper is to know whether the lack of the perfective/imperfect contrast concerning the past tenses in English has an effect in the way machines handle imperfective readings in their Spanish translations. The second question is to know whether human translators are more sensitive to imperfective readings compared to machines. Regarding the first question, our results shown that imperfective readings were significantly less frequent in the translated outputs provided by machines, compared to humans. Machine learning approaches support the idea that the better the MT system is, the more its output will resemble to human-generated text (Corston-Oliver et al. 2001). Our data showed convergent results from human translators and MT outputs in the use of the passive voice in Spanish, which was not frequently used. In turn, our data revealed that the perfective/imperfect contrast was poorly realized by MT compared to humans. Our results account for the weakness of data-driven evaluation methods which, accordingly to Corston-Oliver et al. (2001), "should include measurements of

idiomaticity and techniques to verify that the semantic and pragmatic content of the source language had been successfully transferred to the target language".

Traditionally, scholars have explored typological differences between languages to test universal mappings between meaning and syntax. We find this approach useful to gain insight in translation studies and more specifically in the cognitive processes which are involved in multilingual tasks. The distinction between the preterit and the imperfect is one of the most difficult features for English-speaking students to grasp (Westfall and Foerster 1996). The qualitative comparison of machine and human translations (cf. Table 1 & 4) exemplifies how the preterite and the imperfect interact to determine discourse dynamics. According to the Discourse Representation Theory, the perfective aspect introduces a new reference time that accounts for the advancement of the narrative line (Kamp 1981: 57). On the other hand, sentences carrying the imperfect do not make the narrative action to advance as they merely invoke an available past reference time. By translating the perfective *get*- passive by a majority of past simple verbal forms, machines are providing a conspicuous narrative line in which every predicate moves the narrative time forward. The low frequencies of imperfective aspectual constructions in machines' segments seem thus to have an effect on the readability of their outputs in the sense that they are lacking static narrative anchors allowing to accommodate a set of specific actions within. Our results suggest that the computing process of Google's machines fail to provide a native-like perfective/imperfective contrast in Spanish.

Because the output segments of machine translations are the result of computational models, its results are likely to be poorer when the system translates something in which it has not been trained (Lotz and Van Rensburg 2016). The machines' deficit concerning the production of

imperfective readings in Spanish translations should decrease whenever the model integrates accurate translated samples provided by humans.

On the other hand, human translators proved to largely preserve imperfective readings compared to machines. However, their translations did not reveal a balanced contrast, and the use of perfective forms still corresponded to 57% of their outputs, against 32% marked by imperfective constructions. We cannot rule out the possibility that the translators of *Dreams of my Father*, *Sula*, and *Fahrenheit 451* had been negatively influenced by the source language's construction. Maintaining perfective constructions as Spanish equivalents of the perfective *get*- passive can be taken as a covert translation strategy[21] inherited from the English morphosyntactic layer. A qualitative analysis of each translation would be necessary in order to elucidate whether the distribution of perfective and imperfective readings by human translators are discursively congruent. Nonetheless, it seems clear that machines cannot pick up one or another the same way humans do. Such a decision seems to be highly context dependent for humans. The results of a study exploring the simple present / present progressive aspectual contrast in Spanish revealed that inherent semantic features influence tense selection, namely the progressive was selected to indicate focus on the durative aspect of activities, while the simple tense served to emphasize the telic and punctual aspectual features of achievements (Mayberry 2011). However, other textual cues such as lexical items may result critical to realize this contrast, for example the presence of certain temporal adverbs. In this domain, Cuza (2010) found out that the selection of imperfective constructions was enforced by the presence of frequency adverbs denoting habitual situations (e.g., *normalmente*, usually).

[21] By the term '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.

Our findings indicate that the realization of the Spanish perfective/imperfective contrast by machines was significantly different from the realization performed by human translators. The latter combined more frequently past simple forms and imperfective constructions, whereas the former shown a reiterative use of past simple forms and a poorer use of imperfective readings. These differences between groups remind those found between native speakers and L2 learners in some SLA studies. It has been claimed that grammatical representations based on L2 grammars may diverge not only from those of the target language's native speakers but also from their L1 monolingual counterparts (Hawkins & Chan 1997). This can be explained from a generativist perspective arguing the Failed Functional Features Hypothesis, which assumes that formal features depending on functional categories that were not selected earlier in life are no longer accessible in adulthood, and are likely to lead to diverse fossilization phenomena[22]. While machine translations are expected to evolve in the next years towards the human-like pattern regarding the realization of the Spanish aspectual contrast, it is uncertain whether human translations will contemporarily become more dependent on online machine outputs. Raising awareness on the finest quality provided by human translators should encourage editors, translation teachers and students to use online machine devices with caution.

Conclusion

The goal of our survey was to test whether and to what extent two prominent properties of Spanish were preserved when translating perfective *get*- passive predicates. The first property

[22] However, empirical evidence against this hypothesis has been demonstrated for a group of English near-native learners of Spanish who managed to overcome the lack of perfective/imperfective contrast of their L1 and performed like native Spanish when confronted to two semantic interpretations judgment tasks (Montrul & Slabakova 2003). The conclusions of these authors seem to encourage the idea that access to the universal grammar do not decay with age even for learners not living in the target language environment. More recently, Gómez Soler (2015) has shown that L2 learners can successfully overcome the poverty-of-stimulus argument by resorting to universal principles that guide the grouping of predicates into semantic classes after two judgment tasks dealing with the recognition of aspectual distinctions in Spanish psych-predicates.

had to do with the realization of undergoer arguments in Spanish by means of reflexive or impersonal active-voice constructions rather than the source passive scheme. This property proved to be preserved both in the human and machine translations. Translating the perfective *get*- passive by active-voice constructions seems congruent with the target language's uses. The second property we analyzed was the perfective/imperfective contrast within the translated segments. This property proved to be dissimilarly performed by humans and machines, the former providing a higher number of imperfective constructions over the latter. Machines failed at accurately transposing the internal constituency of an action or state as [± perfective]. Our results suggest that machine translation struggles with two phenomena widely observed in linguistics –i.e., natural languages' richness and individual differences within groups.

Acknowledgments

I express my gratitude to the chief editor and anonymous reviewers for their productive insights. Any errors that remain are mine.

References

Agnieszka, K. (2012). *English and Spanish passive voice in the written media: A contrastive study*. Dissertation/Thesis. Universidade da Coruña, Departamento de filoloxía inglesa.

Anderwald, L. (2018). Language change and cultural change: The grammaticalization of the *get*- passive in context. *Language & Communication* 62, 1-14.

Ayoun, D. (2013). *The Second Language Acquisition of French Tense, Aspect, Mood and Modality*. Amsterdam: John Benjamins.

Babych, B., Hartley, A. (2004). Modelling legitimate translation variation for automatic evaluation of MT quality. In *Proceedings of LREC* vol. 3, pp. 833-836.

Bennazo, S., Andorno, C. (2017). Is it really easier to acquire a closely related language? A study on the expression of iteration and continuation in L2 French. In Martin Howard and Pascale Leclercq (Eds.), *Tense-Aspect-Modality in a Second Language. Contemporary perspectives*. Amsterdam: John Benjamins.

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.

Corston-Oliver, S., Gamon, M., Brockett, C. (2001). A machine learning approach to the automatic evaluation of machine translation. In *Proceedings of the 3th Annual Meeting Association for Computational Linguistics*, pp. 148-155.

Cuza, A. (2010). The L2 acquisition of aspectual properties in Spanish. *Canadian Journal of Linguistics / Revue canadienne de linguistique* 55 (2), 181–208.

Dosi, I., Papadopoulou, D., Tsimpli, I. M. (2017). Issues in the acquisition of grammatical aspect in Greek-English bilingual children. In Martin Howard and Pascale Leclercq (Eds.), *Tense-Aspect-Modality in a Second Language. Contemporary perspectives*. Amsterdam: John Benjamins

El-Haj, M., Rayson, P., and Hall, D. (2014). Language independent evaluation of translation style and consistency: Comparing human and machine translations of Camus' novel "The Stranger". In Sojka, P., Horák, A., Kopeĉek, I., and Pala, K., editors, *Proceedings of the 17th International Conference TSD*, volume 8655 of Lecture Notes in Computer Science, Brno, Czech Republic. Gauchola, R. (2012). Une approche typologique fonctionnelle de la voix passive en français et en espagnol. *Lidil* 46, 153-173.

Glushchuk-Oleia, G. (2014). Negación en el sistema de las perífrasis verbales causativas del español actual. *Logos: Revista de Lingüística, Filosofía y Literatura* 24(2), 139-151.

Gómez Soler, I. (2015). The logical problem of second language acquisition of argument structure: Recognizing aspectual distinctions in Spanish psych- predicates. *International Journal of Bilingualism* 19 (6), 627–645.

Hawkins, R., Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research* 13 (3), 187-226.

Kamp, H. (1981). Événements, représentations discursives et référence temporelle. *Langages* 64, 39-64.

Leonard, L. B., Deevy, P., Miller, C. A., Rauf, L., Charest, M., Kurtz, R. (2003). Surface Forms and Grammatical Functions: Past Tense and Passive Participle Use by Children With Specific Language Impairment. *Journal of Speech, Language & Hearing Research* 46 (1), 43-51.

López-Ortega, N. R. (2000). Tense, Aspect and Narrative Structure in Spanish as a Second Language. *Hispania* 38 (3), 488-502.

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

Mayberry, M. (2011). Synchronous Narratives in Spanish: The Simple Present/Present Progressive Aspectual Contrast. *Hispania* 94 (3), 462-482.

Montrul, S., Slabakova, R. (2003). Competence similarities between native and near-native speakers. An investigation of the Preterite-Imperfect Constrast in Spanish. *Studies in Second Language Acquisition* 25 (3), 351–398.

Papineni, K., Roukos, S., Ward, T., Zhu, W-J. (2002). BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for the Computational Linguistics* (ACL), Philadelphia, pp. 311-318.

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

Schwarz, S. (2017). "Like getting nibbled to death by a duck". Grammaticalization of the GET- passive in the TIME Magazine Corpus. *English World-Wide* 38:3, 305-335.

Schwarz, S. (2018). *Passive voices. Be-, get- and prepositional passives in recent American English*. Dissertation/Thesis. Uppsala Universitet, Department of English.

Smith, C. S. (2012). *The Parameter of Aspect*. Studies in Linguistics and Philosophy 43. Berlin: Springer.

Thompson, D., Ling, S. P., Myachykov, A., Ferreira, F., Scheepers, C. (2013). Patient-related constraints on get- and be- passive uses in English: evidence from paraphrasing. *Frontiers in Psychology* 4:848. doi: 10.3389/fpsyg.2013.00848

Toury, G. (1995). *Descriptive Translation Studies and beyond*. John Benjamins, Amsterdam / Philadelphia.

Van Valin, R. D., LaPolla, R. J. (1997). *Syntax. Structure, meaning and function*. Cambridge: Cambridge University Press.

Volansky, V., Ordan, N., and Wintner, S. (2011). More human or more translated? Original texts vs. human and machine translations. In *Proceedings of the 11th Bar-Ilan Symposium on the Foundations of AI* With ISCOL.

Westfall, R., Foerster, S. (1996). Beyond Aspect: New Strategies for Teaching the Preterite and the Imperfect. *Hispania* 79 (3), 550-560.

Appendix

	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.
EB	Brontë, E. (1847, 1996). *Wuthering Heights.* Illinois: Project Gutenberg Etext.	*Cumbres borrascosas.* (2003). Argentina: Biblioteca virtual universal.

GO	Orwell, G. (1949). *1984*.	*1984*. (1980). Traducción de Rafael Vázquez Zamora. Barcelona: Salvat.
JA	Austen, J. (1813, 2008). *Pride and prejudice*. Illinois: Project Gutenberg Ebook.	*Orgullo y prejuicio.* (2006). Traducción de Patricia Franco Lommers. Madrid: Edimat.
MY	Yousafzai, M., Lamb, C. (2013). *I Am Malala*. London: Orion.	*Yo soy Malala* (2013). Traducción de Julia Fernández. Madrid: Alianza.
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	Scott Fitzgerald, F. (1925). *The Great Gatsby.* Australia: Free eBooks.	*El gran Gatsby.* (2014). Costa Rica: Editorial digital.
TM	Morrison, T. (1987). *Beloved*. New York: Alfred A. Knopf.	*Beloved.* (1988). Traducción de Iris Menéndez Sallés. Barcelona: Ediciones B.
TM_2	Morrison, T. (1973). *Sula*. New York: Knopf.	*Sula.* (2001). Traducción de Mireia Bofill. Barcelona: Plaza y Janés.