

---

# Vocabulary Accuracy of Statistical Machine Translation in the Legal Context

**Jeffrey Killman**

jkillman@uncc.edu

Department of Languages and Culture Studies, University of North Carolina at Charlotte, USA

---

## Abstract

This paper examines the accuracy of free online SMT output provided by Google Translate (GT) in the difficult context of legal translation. The paper analyzes English machine translations produced by GT for a large sample of Spanish legal vocabulary items that originate from a voluminous text of judgment summaries produced by the Supreme Court of Spain. Prior to this study, this same text was translated into English but without MT and it was found that the majority of the translation solutions that were chosen for the said vocabulary items could be hand-selected from mostly EU databases with versions in English and Spanish. The paper argues that MT in the legal translation context should be worthwhile if the output can consistently provide a reasonable amount of accurate translations of the types of vocabulary items translators in this context often have to do research on before being able to effectively translate them. Much of the currently available translated text used to train SMT comes from international organizations, such as the EU and the UN which often write about legal matters. Moreover, SMT can use the immediate co-text of vocabulary items as a way of attempting to identify correct translations in its database.

## 1. Introduction

Legal translation is often considered one of the most challenging areas of human translation practice. According to Alcaraz and Hughes, “Probably the greatest single difficulty encountered initially by legal translators is the unfamiliarity of the vocabulary characteristic of this type of discourse” (2002, p. 16), while the second major source of difficulty is the peculiarity of the morphology and syntax (2002, p. 18). Some traditional advice “is to trust nothing, to suspect everything, to check all terms in reliable dictionaries and to develop a close familiarity with the language of the law by constant and careful reading in both languages” (Alcaraz and Hughes, 2002, p. 43). Machine translation (MT) has, to say the least, typically not been recommended in the legal translation context. A perceived high risk of getting things wrong has likely continued to compel legal translators to continue to rely on more time-tested traditional approaches, while overlooking the recent technological improvements which have been made to MT, in particular statistical machine translation (SMT), and the particular aspects which might now make it a worthwhile tool in the legal translation context. This paper investigates how the output of SMT might benefit the translator of a difficult legal text. According to Forcada, “What one needs is to identify the contexts in which one can use MT effectively and to know what can be expected of it” (2010, p. 215). This paper simply seeks to identify what can be expected of MT in the legal translation context.

The specific MT system this paper looks at is the ubiquitous free online (open-domain) SMT system Google Translate (GT). Until somewhat recently GT was mostly reserved for getting the gist of a website or document in a foreign language. It is now being taken more

seriously. GT appears in some of the leading-edge technologies of translation memories (such as the newer versions of SDL Trados and Déjà Vu) to immediately provide translators with the option to post-edit its output when there are no matches in the translation memory databases themselves. It is thought that this might potentially help translators, depending on their needs or preferences. For instance, translators might save time and effort, in typing or reflecting on the translation of a piece of source text or in having to do research online or elsewhere to come up with it. Translators are not always provided a translation solution by a client or by means of a particular set of past translations or a particular translation tool (e.g. a custom-built MT system or termbase) they might be instructed to use in a professional setting. They might have to come up with solutions on their own, and so it is thought that MT output might bear some of the burden.

What if post-editing GT output at the very least could provide translators accurate translations of vocabulary items in a legal text they might have difficulty translating? This might be a modest yet indeed helpful expectation to have for the quality of the output in the context of legal translation, where one often has to spend a good deal of time doing research on such items (Cao, 2007; Biel, 2008; Monzó, 2008). True, support in the areas of morphology and syntax, the second major source of difficulty according to Alcaraz and Hughes (2002), would be most welcome too, but sentences in legal documents tend to have complex structures and most MT systems, whether rule-based or knowledge-based, handle sentences better when they are simple. SMT, in particular, has developed a good reputation for producing good word and short phrase translations, while performing not so well when it comes to overall sentence grammaticality. Moreover, much of the currently available translated text from which SMT statistically draws its translation output comes from international organizations that often write about legal matters, such as the EU and the UN (Koehn 2010, p. 53). On the basis of these premises, this paper seeks to capitalize on what SMT might be particularly good at and what human translators in the context of legal translation are particularly known for experiencing difficulty with (Alcaraz and Hughes, 2002; Cao, 2007; Biel, 2008; Monzó, 2008) and might welcome technological support with.

The paper places its focus on the accuracy of the terminological and phraseological translation choices provided in English by GT for a selection of 621 varyingly difficult vocabulary items included in a Spanish national legal text of judgment summaries: the Civil Division (Sala de lo Civil/Sala Primera) section of the *Crónica de la Jurisprudencia del Tribunal Supremo: 2005-2006* (Reports of Cases before the Supreme Court: 2005-2006)<sup>1</sup>. The *Crónica de la Jurisprudencia del Tribunal Supremo* is “the work which is drafted under the same name annually by authorities (Gabinete Técnico) of the Supreme Court and intended to disseminate the judgments which for various reasons can be considered to be particularly relevant or generate wider interest” (Consejo General del Poder Judicial, n.d., translation ours).

I was involved in the translation of the Civil Division section into English in 2006-2007, which was a part of the coursework included in a legal translation certificate course at the University of Castile-La Mancha in Spain. I found in the majority of cases that consulting multilingual EU resources was especially useful in order to efficiently translate the majority of the vocabulary items that were found to be difficult. In the present study, I will compare machine translations of these items as rendered by the ubiquitous GT, to test the extent to which this popular free online SMT system might translate correctly legal vocabulary that one might otherwise have to spend time and effort on researching.

I will start by providing a brief overview of the recent research on post-editing with GT. I will then discuss in detail the vocabulary items under study, how we approached their transla-

---

<sup>1</sup> The author was involved in the translation of a thousand or so pages of both this version and the 2004-2005 version, which was published in 2006 by Spain’s General Council of the Judiciary.

tion back in 2006-2007, and why SMT is likely able to translate them correctly, before providing details on the methodology. To conclude, I will discuss the results of the test and the implications of the findings.

## 2. Recent GT Research

According to Pym, "Recent research (Pym 2009, Garcia 2010, Lee and Liao 2011) indicates that, for Chinese-English translation and other language pairs, statistical MT [in particular GT] is now at a level where beginners and Masters-level students with minimal technological training can use it to attain productivity and quality that is comparable with fully human translation, and any gains should then increase with repeated use" (2013, p. 2). While none of these studies includes the English-Spanish language pair, "it is reasonable to assume that similar results might emerge in tests using the FIGS languages (French, Italian, German, Spanish), for which data (parallel texts, grammatical rules) have been collected over a longer period of time" (Garcia, 2010, p. 18).

Two of these three recent studies (Pym, 2009 and Lee and Liao, 2011) report on GT's vocabulary accuracy. One of the groups in Pym's study reported appreciating some of the terminology proposed by GT (p. 141). Lee and Liao (2011) include significantly more information than Pym (2009) about vocabulary accuracy in their study. In particular, they find that in the linguistically weaker of the two groups of students that were subjects of their study, "the more words from the MT text a student uses, using [the] sentence as a unit, the less likely a student would make a mistake in translating that particular sentence" (p.128). They report "that the students recognize they can use the MT directly if the meaning is intact, and they would only have to do a little tweaking" (p.128). They note a number of instances where GT translated more accurately than students the contextually appropriate meaning of an ambiguous piece of language (pp.136-137), as well as the specific meanings of adjectives and nouns found in collocations (pp.137-138). As regards the appropriate wording of concepts, they even find that "The divergence in register between the With MT and No MT students was evident" (p. 133).

These studies touch upon the potential of SMT as a viable source of vocabulary translation support. The present study seeks to determine exclusively how accurate SMT is at translating legal vocabulary translators often have to research. It compiles significant empirical data in this very important area.

A focus on lexical data is more practical than one on the morphology and syntax of machine translations of the judgment summaries contained in the Civil Division text. In the majority of cases judgment summaries are drafted in sentences that are unusually long and complex, with features such as multiple subordination and postponement of the main verb until very late in the sentence. Long, complex sentences (which are also common in many other legal text types such as statute laws, judicial rulings, and regulations) will likely not be translated correctly by SMT, or any other type of MT for that matter. What SMT basically does is weigh its decisions on the statistical patterns of various combinations of words found in a sentence and depending on the particular system, the technology might do so with or without traditional MT methods using grammatical rules (in the form of classifications and groupings). While SMT might translate particularly well the vocabulary items of a long sentence by using the context of directly neighboring words, it might translate not so well when it comes to overall grammaticality (e.g. long-distance grammatical problems, unseen morphological forms, etc.). Problems with GT misreading syntax, for instance, were noted by all the translators in Pym (2009), who "were generally appalled by the resulting wild mistranslations" (Pym 2009, p. 141).

To illustrate what can happen to the overall grammaticality of a statistical machine translation of a relatively syntactically complex sentence, Table 1 contains in the first row a single-sentence judgment summary taken from the Civil Division text. The second row contains an unedited Google translation of the judgment summary and the last row, a human-edited version we rendered by using as many correct words and phrases as possible from the Google translation. The errors pointed out result from the syntactic complexity of this single-sentence judgment summary.

---

**Source text**

6.- Derecho marítimo

6.1. La STS 28-9-2005 (RC 769/2005) destaca porque en ella, al examinar un supuesto de responsabilidad por abordaje, diferenciando sus distintas clases, se declara que, sin perjuicio de que las disposiciones contenidas en el Convenio de Bruselas de 23 de noviembre de 1910 sobre unificación de ciertas reglas en materia del abordaje, formen parte del ordenamiento jurídico español y sean de aplicación directa, resulta aplicable la legislación interna, con exclusión de cualquier otra, cuando los buques implicados son de nacionalidad española y el abordaje ha tenido lugar en aguas jurisdiccionales españolas.

**Google Translate**

6. - Maritime Law

6.1. The STS 28.09.2005 (RC 769/2005) stands out because in it, to consider a theory of liability for collision, differentiating their various classes, states that, without prejudice to the provisions of the Brussels Convention of 23 November 1910 on the unification of certain rules relating to the collision, part of the Spanish legal system and have a direct, domestic law applies to the exclusion of any other, when the vessels involved are of Spanish nationality and the approach has taken place in Spanish waters.

**Human-edited version**

6.- Maritime Law

6.1. The Judgment of the Supreme Court of 28-9-2005 (Appeal 769/2005) stands out because when it considers liability for collision by differentiating its various classes, it states that even though the provisions in the Brussels Convention of 23 November 1910 for the Unification of Certain Rules of Law with Respect to Collision Between Vessels are part of the Spanish legal system and are directly applicable, domestic law applies, to the exclusion of any other law when the vessels involved are of Spanish nationality and the collision has taken place in Spanish waters.

---

Table 1. Example of a Civil Division judgment summary accompanied by a Google Translate machine translation and a human-edited version. Highlighted passages indicate the source text grammatical and morphological areas that GT had problems with in the first row, the corresponding Google translations in the second row, and our human solutions and/or modifications in the human-edited version in the third row.

While the overall grammaticality of the GT output is mostly inadequate, most of the vocabulary items and the grammar of individual phrasal items are indeed correct, seeing how many such items could be left unchanged in our human-edited version. An interesting error, however, occurred with the lexically ambiguous word *abordaje*. It was translated correctly as 'collision' (the specialized meaning) the first time it appears and incorrectly as 'approach' (the everyday meaning) the second time it appears. Often, vocabulary in the everyday world takes on another meaning in the field of law. This vocabulary is often referred to as 'semi-technical', one of the areas of legal vocabulary we will describe in the following section.

Garcia (2010) is the only study that includes a text on legal topics<sup>2</sup>. GT might be acceptably accurate for other legal texts, although our text, unlike those in the other studies, is aimed at expert readers and much longer in length with 12, 263 words vs. an average of 199 words. Lee and Liao suggest for the future that “text genres of longer length may provide more in-

---

<sup>2</sup> The text used in Lee and Liao (2011) was a cellphone care instruction guide and the text in Pym (2009), a publication in an online dictionary of technical terms.

depth insights into MT use” (2011, p. 143). Our long text provides a sizeable sample of 621 legal vocabulary items on which this paper performs an in-depth empirical analysis of MT accuracy. To our knowledge, no such analysis has been performed yet.

### 3. Nature of the Vocabulary Items under Study

614 of the 621 vocabulary items under study can be classified as symbolic items, while the remaining 7 can be classified as functional items. Functional items are:

grammatical words or phrases that have no direct referents either in reality or in the universe of concepts, but which serve to bind together and order those that do. Examples from the legal sphere are 'subject to', 'inasmuch as', 'hereinafter', 'whereas', 'concerning', 'under' and 'in view of'. Deictics, articles, auxiliaries, modals and other purely syntactic and morphological markers also belong with this group, as do other more complex units like 'unless otherwise stated', 'as in section 2 above', 'in accordance with order 14' and similar phrases (Alcaraz and Hughes, 2002, p. 16).

The 7 functional items are varyingly complex conjunctions or prepositional phrases such as *al régimen de* (under), *en atención a* (in view of), *según lo dispuesto en* (pursuant to), *ex artículo* (referred to in article). Certain functional items may have a high level of frequency in legal texts or be somewhat peculiar to them. Either way, the translator must be aware of them and know what their most contextually appropriate equivalents are in the target language. For example, 'pursuant to' is a frequent pattern among legal texts and, by extension, arguably characteristic of the discourse. In everyday language one would likely use 'according to' instead. In most cases, the individual units of multiword functional items cannot be translated one-to-one. The items are not compositional and equivalence must be established by the translator on the basis of probably all the co-occurrences in the phrase. For example, 'to the regime of', a literal word-for-word translation of *al régimen de*, would not make sense as a substitution for 'under' in 'under the contract', a possible translation of *al régimen del contrato*.

The 614 symbolic items "refer to things or ideas found in the world of reality, physical or mental" (Alcaraz and Hughes 2002: 16). 421 of them can be classified as terms and 193, as phrases. A term is defined as "a designation of a defined concept in a special language by a linguistic expression" and "can consist of single words or be composed of multiword strings. The distinguishing characteristic of a term is that it is assigned to a single concept, as opposed to a phrase, which combines more than one concept in a lexicalized fashion to express complex situations" (ttt.org-CSL Framework, 2001, ISO 12620 Data Categories). For example, *conocimiento de embarque* (bill of lading) is a term, whereas *bajo el régimen de conocimiento de embarque* (under a number of bills of lading) is a phrase.

The 614 symbolic items (421 terms, 193 phrases) can be divided into five vocabulary groups: purely technical vocabulary (221 terms, 110 phrases), semi-technical vocabulary (100 terms, 26 phrases), everyday vocabulary frequently found in legal texts (62 terms, 56 phrases), and official legal vocabulary (38 terms, 1 phrase). The purely technical, semi-technical, and official items can be classified under the following areas of law depending on the case: procedural law (148 cases), civil law (144 cases), commercial law (123 cases), constitutional law (18 cases), family law (12 cases), criminal law (10 cases), international law (7 cases), tax law (8 cases), European Union law (7 cases), administrative law (6 cases), inheritance law (5 cases), insurance law (5 cases), employment law (2 cases), and United Nations law (1 case).

In the following paragraphs each of the five types of vocabulary items are described in more detail.

Technical items, “whatever their origins may have been, now belong almost exclusively to the vocabulary of the law, or are firmly attached to this sphere in their everyday usage” (Alcaraz and Hughes, 2002, p. 154). Single-word examples (and their English translations) taken from the Civil Division text (and its English translation) include *demandado* (defendant), *exequátur* (exequatur), *homicidio* (manslaughter), *litigio* (litigation), *testamento* (will), *testador* (testator). While each of these items mostly has a single meaning, others “are more complex in that their precise senses may vary with context, though the overall field of reference remains the legal one” (ibid, p. 156). For example, *jurisprudencia* may be translated as ‘jurisprudence’ (if referring to the science of the law) or ‘case law’ (if referring to the decisions of judges relating to particular matters in contrast to statute law), and *embargo* as ‘seizure’ (in the context of civil law) or ‘embargo’ (in the context of international law). Technical multi-word terms and phrases “are meaningful only in a legal context, even though the individual words of which they are composed may belong to the general vocabulary of everyday speech” (ibid, p. 157). The following are multiword examples (and their English translations) taken from the Civil Division text (and its English translation): *cuestiones jurídicas suscitadas* (points of law that arise), *derechos reales* (rights in rem), *desestimación de la acción* (failure of the action), *desestimar el recurso interpuesto por* (dismiss the appeal brought by), *ejecución de una hipoteca* (foreclosure), *modificación jurisprudencial* (departure from precedent), *reparación íntegra del daño* (full compensation for damage), *resolución del contrato* (termination of contract). Some of the individual words of these multiword units may even have other meanings in other contexts. For example, other potential contextual meanings of *reales*, *reparación*, *recurso*, or *resolución* could be ‘real’, ‘reparation’, ‘resource’, or ‘resolution’, respectively. Some of these multiword items may only be properly understood by considering all the co-occurrences together and hence cannot be translated using a word-for-word approach. For example, *ejecución de una hipoteca* and *modificación jurisprudencial* would not make sense if rendered as ‘execution of a mortgage’ and ‘jurisprudential modification’, respectively.

Semi-technical vocabulary “is a more complex group, since it contains terms [and phrases] that have one meaning (or more than one) in the everyday world and another in the field of law” (ibid, p. 158-159). The items “belonging to this group are more difficult to recognize and assimilate than wholly technical terms [and phrases]” (ibid, p. 17). A phrase example (and its English translation) taken from the Civil Division text (and its English translation) is *causas de* (grounds for). Outside the sphere of procedural law it could be rendered as ‘causes of’. Another phrase example is *beneficios derivados de* (profits arising from), which could be rendered as ‘benefits deriving from’ outside of the sphere of commercial law. A multiword term example is *masa activa* (insolvency estate), which could be translated as ‘active mass’ outside the domain of commercial law. Another multiword term is *acción infundada* (unmeritorious proceedings), which could be translated as ‘unfounded action’ outside the domain of procedural law. A few single-word term examples are *prenda* (pledge), *sociedades* (corporations), and *competencia* (jurisdiction), which in the everyday world could be translated as ‘garment’ (piece of clothing), ‘societies’, and ‘competition’, respectively.

Everyday vocabulary frequently found in legal texts consists of items “in general use that are regularly found in legal texts but, unlike the previous group, have neither lost their everyday meanings nor acquired others by contact with the specialist medium” (Alcaraz and Hughes, 2002, p. 18). Translators “will probably find that the terms [and phrases] are easier to understand than to translate, precisely because they tend to be contextually bound” (ibid: 162). A phrase example (and its English translation) taken from the Civil Division text (and its English translation) is *realidad y efectividad* (tangible effect). While each of the words in the phrase is being used in the most general sense, the literal rendering in English of each of them ‘reality and effectiveness’ might be awkward or nonsensical to the reader, and so it may

well be difficult for the translator to come up with an appropriately worded translation even though the source phrase is not difficult to comprehend. There are also everyday items that may have more than one everyday meaning depending on the context. A single-word term example is *omisión*, which in the Civil text document was translated as ‘failing to act’, not ‘omission’, which is how *omisión* would be translated were it being used according to its most common everyday meaning. The translator might not recall right away the less common meaning and might not be able to quickly produce a contextually appropriate translation. There are also everyday items that while they may not have a contextually sensitive meaning or wording, their register is high or their use is uncommon. A few examples of high-register items (and their English translations) taken from the Civil Division text (and its English translation) are: *homologar* (approve) instead of, say, *aprobar*; *decantarse por* (opt for) instead of, say, *favorecer*; and *coadyuvar* (contribute) instead of, say, *contribuir*. A few examples of not-commonly-used items (and their English translations) taken from the Civil Division text (and its English translation) are: *pleno* (plenary session), *su vigencia* (its period of operation), and *información veraz* (honest information).

Official legal vocabulary can include the names of specific laws, conventions, titles of legal professions or documents, etc. Official items that are part of a supranational entity often have official equivalents in other languages. The translator must already know or find out what they are. Some term examples (and their English translations) taken from the Civil Division text (and its English translation) are *Artículo 17 del Convenio de Bruselas de 27 de septiembre de 1968* (Article 17 of the Brussels Convention of 27 September 1968), *Convención de las Naciones Unidas sobre Contratos de Compraventa Internacional de Mercaderías, de 11 de abril de 1980* (United Nations Convention on Contracts for the International Sale of Goods of 11 April 1980), *Principios de Derecho Europeo de Contratos* (Principles of European Contract Law). While the official items that are part of the national legal system of Spain do not have official equivalents in English, there might be translations available in English that are time-tested and/or documented in important dictionaries or databases and thus often used by translators. Even so, English translations of the official vocabulary of the legal system of Spain do not have to be worded in any specific way as long as they are semantically correct and not "clumsy or too long to be practical" (Biel 2008, p. 34). Some examples (and their English translations) from the Civil Division text (and its English translation) are *Boletín Oficial de la Provincia* (Provincial Official Gazette), *Ley 27/1992, de Puertos del Estado y de la Marina Mercante* (Law No 27/1992 on National Ports and the Merchant Navy of 24 November 1992), *Texto Refundido de la Ley de Sociedades Anónimas* (consolidated version of the Law on public limited companies). Like the other categories, they may also include ambiguous words.

What particularly stands out is a high level of contextual determinacy either in how the source language items are meant to be understood or in how they should actually be written in the translation. This 'contextual sensitivity' seems a large part of what compels legal translators to spend time looking vocabulary items up in other resources, to verify how a source language item should be interpreted or how it should be written in the translation. The following brief paragraphs summarize what is specifically meant by 'contextually sensitive' as it pertains to each group of vocabulary items.

The largest group in which all the vocabulary items are contextually sensitive is that of semi-technical vocabulary (100 terms, 26 phrases). These items are grouped together because they have at least two different contextual meanings, a legal one and an everyday one, and, in theory, may be interpreted in more than one way.

The other group in which all the items are contextually sensitive is that of the 7 functional phrases, which may either need to be translated in a legally peculiar way or may only be properly understood on the basis of all the co-occurents.

The 331 purely technical vocabulary items are contextually sensitive in 163 cases (98 terms, 65 phrases) either because single-word terms may have more than one legal sense or because multiword terms or phrases may contain ambiguous words or the multiword units themselves may only be properly understood on the basis of all the co-occurents.

The 118 everyday items frequently found in legal texts are contextually sensitive in 51 cases (26 terms, 25 phrases). The category consists of single-word terms that may have more than one sense, multiword terms or phrases that contain ambiguous words or the multiword units themselves may only be properly understood on the basis of all the co-occurents, as well as items whose translation needs to be worded in a specific manner. An example of a phrase that needed to be translated in a particular way is *comportamiento del administrador* (conduct of the administrator). *Comportamiento* in most cases is translated as ‘behavior’, but since the situation refers to one’s behavior in a professional capacity, ‘conduct’ is preferred.

The 39 official vocabulary items are contextually sensitive in 23 cases (22 terms, 1 phrase) because either they have official equivalents in English that need to be worded in a specific manner (13 cases) or contain ambiguous words (7 cases).

On the basis of these parameters, context in 370 out of the 621 total cases (i.e., 60% of the cases) was assessed to be critical in how a source language item should be interpreted or in how its translation should be written.

#### 4. Background: Documentation Experience

When in 2006-2007 we translated the Civil Division text, we prioritized using online EU multilingual databases containing English and Spanish language versions over any other non-EU resource when we were faced with translating the difficult vocabulary under study. Our main reasons for doing so were: 1.) to test how applicable EU resources could potentially be for translating the vocabulary found in a legal text produced by the Supreme Court of Spain (which is a member of the EU), 2.) because we wanted our English to be worded as close as possible to that used at the EU level, as the translation specifications did not include any particular English-speaking target culture and the translation was in theory being done for European readers in general (not necessarily native speakers of English), and 3.) for convenience purposes—EU resources are available free of cost and online.

595 (i.e. 96%) of the 621 vocabulary items were found to have a suitable equivalent in existing free online EU multilingual databases, such as EUR-Lex (Access to European Law) in 345 instances, IATE (inter-active terminology for Europe) in 247 instances, and the European Judicial Network (EJN) in 3 instances<sup>3</sup>. The other 26 cases could in 17 instances be documented in Alcaraz and Hughes’s Spanish-English *Diccionario de Términos Jurídicos* (2005) and in the remaining other 9 instances, in various other non-EU resources.

In all cases we documented the longest piece of text possible. This was preferable because longer items include more context and consequently more information about how they

---

<sup>3</sup> EUR-Lex and IATE are the two multilingual resources that are used most by the translators of the largest translation organization in the world: the Directorate-General for Translation of the European Commission (European Commission, 2009, pp. 14-15). The multilingual corpus EUR-Lex, available online since 2002 (Liebwald, 2009, p. 274), provides free access to European Union law and other documents considered to be public. Released officially to the public in March 2007, the terminology database IATE “is a shared and interactive term base of the institutions and other bodies of the EU, designed as the main instrument for the multilingual drafting of Community texts” (Muñoz and Valdivieso, 2009, p. 375, translation ours). IATE and its beta version released in 2006 were available for use at the same time the translation of the Civil Division text was being done and during the editing phase. The EJN, for its part, contains information about Member States, their civil and commercial law, and European law.



need to be translated. As pointed out in the previous section, context in 60% of the cases was critical in how a source language item should be interpreted or in how its translation should be written. In any event, 463 out of the 621 cases (75%) contained more than one word (i.e., 200 phrases and 263 multiword terms). Terminological resources such as IATE, in the case of the EU resources, or Alcaraz and Hughes (2005), in the case of the non-EU resources, were consulted first, because they are more user-friendly. While a multilingual document repository like EUR-Lex might offer many possibilities (including terms and phrases that are not registered in a specific terminological resource), one can end up spending too much time sorting through too many parallel documents in its database. So it might be more efficient to first try using a high quality terminological resource, if one is indeed available (see Melby [2012] for an in-depth discussion of the role and importance of high quality term bases in the age of aligned multilingual corpora). Nevertheless, no matter how efficiently one tries to go about it, a good deal of time can be spent gathering, inputting, or sifting through written data.

## **5. Guarded Optimism**

There are reasons to believe that SMT, "currently the dominant paradigm in MT research [with] a growing share of the MT market" (Forcada, 2010, p. 221), should be able to provide suitable translations for our sample of 621 researched terms and phrases. On the one hand, much of the currently available translated text used to train SMT comes from international organizations that often write about legal matters, such as the EU and the United Nations (Koehn, 2010, p. 53). On the other, SMT can use the immediate co-text of terms and phrases to identify translations that are likely correct for potentially ambiguous pieces of text whose meanings are hinged on the other words which surround them. SMT draws its solutions from an enormous database of previously translated texts where millions of sentences in one language have been aligned with their human translations in the other language. If provided the option of a longer phrase translation, SMT tends to use it (as we set out to do when documenting translation solutions back in 2006-2007), although longer phrases are less frequent and hence less statistically reliable (Koehn, 2010, p. 141). Even though a longer phrase translation is there, SMT might not select it, as there may not be enough statistics and hence enough reliable probability estimates.

In any event, the main hypothesis was that more often than not GT would be able to provide suitable translations for the 621 vocabulary items. There were also two sub-hypothesis: 1.) GT would in a good number of cases be able to provide suitable translations for 'contextually sensitive' pieces of text thanks to its phrase-based statistical approach. 2.) At the same time however, most of the errors would likely be produced when translating 'contextually sensitive' words or phrases. On the one hand, disambiguating ambiguous words and phrases has long been the Achilles' heel of MT ever since its first generation of systems. MT, however, has greatly improved thanks to the statistics of frequencies and corresponding probabilities drawing from a potentially comprehensive corpus of past translations and its ability to inform its decisions by using immediate co-text. On the other hand, there is no guarantee that the translation solutions provided by the open-domain GT will be worded in a desirable way. The database may not contain translations with a desired wording or the system may not heuristically be able to draw on them.

## **6. Methodology**

While it would have been quicker for analysis purposes to feed only the vocabulary items under study into GT, the results in many cases would have likely not benefitted from access to

the surrounding co-text the system might be able to use as relevant input. Therefore, we fed the entire source text into GT, but one judgment summary section at a time for ease of data collection. A difference would not have been made in the MT results had the entire source text been fed into GT all at once. While SMT may on occasion be able to source from its database a translation for a previously seen sentence, its recognition of co-text does not stretch beyond the sentence, which is why it is said that MT translates a text “as a bag of unconnected sentences” (Hardmeier and Federico, 2010, p. 288).

Machine translations that were different from the handpicked ones were analyzed to determine whether they were correct. Though absolute synonymy in a given language is often thought not to be possible, translators may often be able to choose among several options to express the same thing in a legal translation (Mayoral, 2004, p. 61).

This study uses “a harsh correctness standard – is the translation perfect or not?” (Koehn, 2010, p. 218). There is a reasonable chance no MT mistakes are made at the term and phrase level. Because “correctness may be too broad a measure”, the study breaks it down into “the two criteria fluency and adequacy”; the former “involves both grammatical correctness and idiomatic word choices”, while the latter asks: “Does the output convey the same meaning as the input [...]? Is part [or all] of the message lost, added, or distorted?” (Koehn, 2010, p. 218). Correct renditions are judged to be both ‘fluent’ and ‘adequate’, while incorrect ones are judged to be either ‘influent’ or ‘inadequate’, or both. While a harsh standard may overlook ‘incorrect’ translations of multiword items that are not entirely incorrect, it discourages the leniency variability that can occur with multiple-range scoring. Moreover, partially incorrect output would in theory not spare one from having to do research, though it might prompt one to devise a more efficient research strategy.

When an item being used in the same way more than once was machine translated differently on more than one occasion, the MT variations were recorded, although the item was classified according to the ‘best’ possible variant. For example, if a particular item was translated incorrectly in one instance but correctly in another, it was counted as correct under either the category of the ‘same’ translations or the ‘different but correct’ ones, depending on the case. If a machine translation of an item was ‘different but correct’ in one instance and the ‘same’ in another, that which was the same was counted<sup>4</sup>. Different source-text contexts and the use of free online SMT on different occasions can prompt different machine translations. When an item appearing more than once was consistently machine translated in the same way, this was not taken stock of<sup>5</sup>.

## 7. Results

223 of all the results (36%) were machine translated ‘in the same way’, 177 (28.5%) were machine translated ‘in a different but correct way’, and 221 (35.5%) were machine translated ‘incorrectly’. That a little over 64% of the results were machine translated correctly supports our main hypothesis that more often than not GT is able to provide suitable translations for the 621 items. In Figure 1 the percentages of these overall results can be compared with those of the results of each of the 5 vocabulary categories under study.

---

<sup>4</sup> Opting for the machine translation that was the same as our previously documented translation solution helped speed up data collection, as ascertaining whether a different machine translation was correct would often entail more investigative work on the part of the researcher. In other words, our hand-picked solutions were not ‘better’ by default than any other possible correct MT rendition; they were merely less burdensome to categorize.

<sup>5</sup> This measure was also less burdensome for data collection purposes.

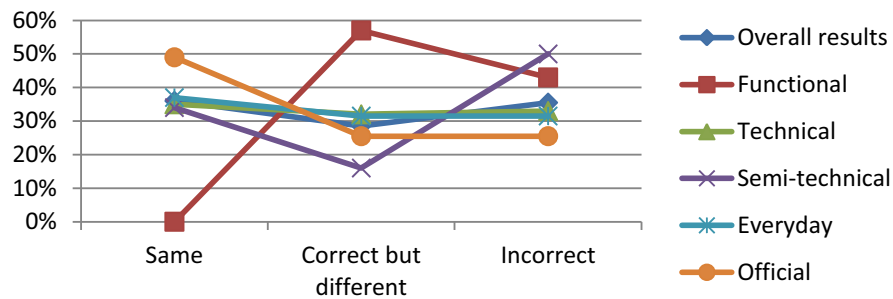


Figure 1. Percentages of the overall results and those of the results of each of the 5 vocabulary categories under study.

The technical and everyday items are the two categories whose results are most consistent with the overall ones. They are significant in volume; the former category, the largest of all the other categories with its 331 items, comprises some 53% of the entire sample under study and the latter category with its 118 items, some 19%. The results of the remaining three categories deviate most from the overall ones. The functional and the official items, due to their limited volume: 7 (1%) and 39 (6%), respectively, might not be the most statistically reliable and quite possibly random. However, the semi-technical items are reasonably significant in volume, with 126 units making up some 20% of the entire sample. Either way, both the semi-technical and the functional items are the two entirely ‘contextually sensitive’ categories and they have the highest error percentages. It is also worth pointing out that in each of the remaining categories (i.e., those in which not all the items are contextually sensitive) the ‘incorrect’ results involved the highest concentrations of contextually sensitive vocabulary. Figure 2 compares the percentages of contextually sensitive items in each area of the results of the remaining vocabulary categories with the overall results.

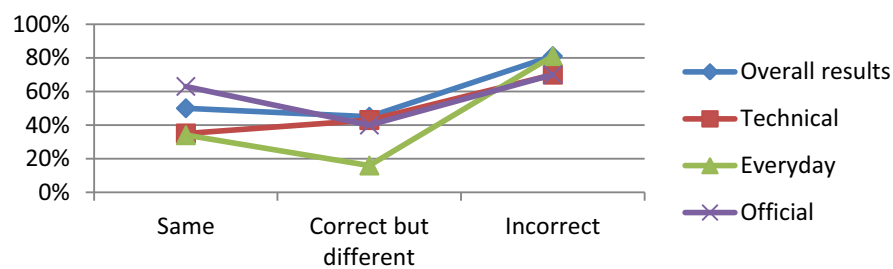


Figure 2. Percentages of ‘contextually sensitive’ items in each area of the overall results and the results of the remaining vocabulary categories.

The overall results in Figure 2 show that in a little over 80% of all the cases in which GT translated incorrectly, contextually sensitive items were involved. This absolutely supports our second sub-hypothesis that most of the errors would likely occur when translating contextually sensitive items. But looking at the 370 total ‘contextually sensitive’ items as a whole (i.e. 60% of the entire vocabulary sample) reveals that 111 of them were machine translated ‘in the same way’, 80 ‘in a different but correct way’, and 179 ‘incorrectly’. That a little over half of them were machine translated correctly supports our first sub-hypothesis that GT would in a good number of cases be able to provide correct translations for ‘contextually sensitive’ items. Accuracy in as many as around half the cases was not expected.

Take as an example *según lo dispuesto en*, a functional item originally documented in EUR-Lex as ‘pursuant to’. GT translated it correctly as ‘under the provisions of’; it remarkably did not grind out a word-for-word error to the likes of ‘according to that available in’. Another example is *propietario y explotador*, a contextually sensitive technical phrase from commercial law that was originally documented in EUR-Lex as ‘owner and operator’ and then machine translated in the same way. In other contexts *explotador* could be translated as ‘exploitative’ or ‘exploiter’, for example. It seems GT was able to use *propietario* as a contextual cue. A couple of error examples, however, are the semi-technical procedural law terms *proceso de ejecución* and *acción ejecutiva*. The first was originally documented in EUR-Lex and the second, in Alcaraz and Hughes (2005). The translation solution arrived at in each case was ‘enforcement proceedings’, as it turns out the terms were being used synonymously. GT mistakenly rendered them according to their non-technical meanings as ‘implementation process’ and ‘executive action’, respectively. More contextually sensitive examples are described in the Appendix.

## 8. Conclusion

The results of this study show that GT could indeed translate accurately vocabulary taken from a voluminous legal text aimed at expert readers in a little over 64% of the cases. Little time is invested in using this free online SMT system. It produces translations for words and phrases instantly. That it is able to do so correctly at least more often than not in an area of translation widely considered one of the most difficult renders it, in our opinion, a considerably reliable tool legal translators might indeed find beneficial. I’ve yet to come across a tool that is (close to) 100% effective, let alone in the legal translation context. SMT may come in especially handy when it is able to correctly translate the difficult vocabulary one might otherwise need to spend time looking up. Manually looking up terms and phrases over and over again can take up a good deal of time and ideally should be kept to a minimum. A reliable SMT system might significantly cut down on the amount of vocabulary one might need to look up.

When we translated without MT in 2006-2007, MT was, to put it mildly, not thought of as a viable source of translation support, let alone for legal texts, often characterized by contextually sensitive vocabulary and complex discourse structures. Context in a little over 80% of the MT errors reported in this paper was indeed a factor, although a little over half of all the ‘contextually sensitive’ cases were machine translated correctly. According to these results, MT’s contextual Achilles’ heel still remains vulnerable, but certainly not entirely at a little less than 50% thanks to recent phrase-based statistical methods. In any event, one must keep in mind that as long as the database contains enough translations that are similar to the text being translated, the less ambiguous the co-text of the input and the likelier it is to be a frozen pattern of language, the likelier it is SMT will translate accurately in terms of correct word-sense disambiguation. The co-text that can be read by the machine will likely be longer and hence more likely to be translated according to its contextually determined meaning.

It is hoped that this paper will shed light on the quality of free online SMT in the legal translation context. Having an idea of the aspects in which it might perform particularly well or bad in a specific context is important so that one may effectively work with the output if one so chooses and set reasonable expectations for it. This paper demonstrates with a sizeable sample of legal vocabulary, considered “the greatest single difficulty encountered initially by legal translators” (Alcaraz and Hughes, 2002, p. 16), that free online SMT might perform consistently well in this area. Legal translators, especially those that are new to the field, might find that SMT definitely has something to offer in terms of vocabulary.

## References

- Alcaraz, E. and Hughes, B. (2002). *Legal Translation Explained*. St. Jerome Publishing, Manchester.
- Alcaraz, E and Hughes, B. (2005). *Diccionario de términos jurídicos*, 8th ed. Editorial Ariel, Barcelona.
- Biel, L. (2008). Legal Terminology in Translation Practice: Dictionaries, Googling or Discussion Forums? *Skase*, 3(1):22-38.
- Cao, D. (2007). *Translating Law*. Multilingual Matters Ltd, New York.
- Consejo General del Poder Judicial. (nd). Crónica de jurisprudencia. [http://www.poderjudicial.es/cgpj/es/Poder\\_Judicial/Tribunal\\_Supremo/Actividad\\_del\\_TS](http://www.poderjudicial.es/cgpj/es/Poder_Judicial/Tribunal_Supremo/Actividad_del_TS)
- Forcada, M. L. (2010). Machine Translation Today. In Gambier, Y. and Doorslaer, L.V. (eds) *Handbook of Translation Studies Volume 1*, pages 215-223, John Benjamins, Amsterdam.
- Garcia, I. (2010). Is Machine Translation Ready Yet?. *Target*, 22(1):7-21.
- Hardmeier, C. and Marcello, F. (2010). Modelling Pronominal Anaphora in Statistical Machine Translation. In *Proceedings from the 7th International Workshop on Spoken Language Translation*, pages 283-289, Paris.
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge University Press, New York.
- Lee, J. and Liao, P. (2011). A Comparative Study of Human Translation and Machine Translation with Post-Editing. *Compilation and Translation Review*, 4(2):105-149.
- Liebwald, D. (2009). Interfacing between Different Legal Systems Using the Examples of N-Lex and EUR-Lex. In Grewendorf, G. and Rathert, M. (eds) *Formal Linguistics and Law*, pages 257-291, Mouton de Gruyter, Berlin.
- Mayoral, R. (2004). Lenguajes de especialidad y traducción especializada: la traducción jurídica. In Gonzalo, C. and Garcia, V. (eds) *Manual de documentación y terminología para la traducción especializada*, pages 49-71, Arco/Libros, Madrid.
- Melby, A. (2012). Terminology in the Age of Multilingual Corpora. *JOSTRANS*, 18:7-29.
- Monzó, E. (2008). Documentación para la traducción inglés-español. In Ortega, E. (ed) *La traducción e interpretación jurídicas en la Unión Europea*, pages 733-752, Comares, Granada.
- Muñoz, M. and Valdivieso, M. (2009). La terminología en las instituciones de la UE: de la fragmentación a la convergencia. *EntreCulturas*, 1:365-383.
- Pym, A. (2009). Using Process Studies in Translator Training: Self-Discovery through Lousy Experiments. In Göpferich, S., Alves, F., and Mees, I.M. (eds) *Methodology, Technology and Innovation in Translation Process Research*, pages 135-156, Samfundslitteratur, Copenhagen.
- Pym, A. (2013). Translation Skill-Sets in a Machine-Translation Age. *Meta*, 58(3):487-503.
- ttt.org-CSL Framework. (2001). ISO 12620 Data Categories. <http://www.ttt.org/clsframe/>

## Appendix. GT Machine Translations of Contextually Sensitive Vocabulary

### Rendered in the Same Way

A semi-technical example is *balance de situación*, a commercial law term that was originally documented in IATE as ‘balance sheet’ and then machine translated in the same way. It is remarkable that it was not machine translated literally and incorrectly as ‘balance of situation/situation balance’.

*Omisión del deber de* is an everyday phrase that was originally documented in EUR-Lex as ‘failure to’ and then machine translated accordingly. An awkward or potentially nonsensical word-for-word translation might have been ‘omission of the duty of’. Another everyday example is *vicio oculto*, a term originally documented in IATE as ‘latent defect’ and then machine translated in the same way. It is noteworthy that GT did not translate the two-word string as ‘hidden vice’, for example, ‘hidden’ and ‘vice’ being appropriate translations of *oculto* and *vicio*, respectively, in most other contexts.

An official term example is the European Union *espacio económico europeo*. Its official equivalent in English: ‘European economic area’, which we documented in IATE, was the MT of GT. Obviously, GT was able to use the co-text of this multiword term and match it to relevant examples existing in its database. A word-for-word incorrect rendition might be “European economic space” or “European financial space”, as “space” and “financial” are often translations of *espacio* and *económico*, respectively.

### Rendered in a Different but Correct Way

A technical term example is *abuso de derecho*. It belongs to procedural law and was originally documented in IATE as ‘misuse of law’ and then machine translated differently but correctly as ‘abuse of process’. In most contexts *derecho* may be translated as ‘right’ or ‘law’. Nevertheless, GT was able to decide on the suitable translation of *derecho* by using the relevant co-text: *abuso de*. Another technical example of a procedural law item containing *derecho* is the phrase *ajustado a derecho*, which was originally documented in EUR-Lex as ‘legally sound’ and then machine translated as ‘consistent with the law’. Had GT rendered *ajustado a* as ‘adjusted to’, which would be correct in the majority of other contexts, the MT would be unidiomatic and hard to understand.

An everyday term whose translation has to be carefully worded is *el elemento determinante*. We produced ‘the decisive factor’ thanks to EUR-Lex. GT, though, came up with ‘the determinant’. Other suitable renditions might be ‘the crucial element’ or ‘the determining factor’. But ‘the determining element’, an easily comprehensible word-for-word rendition, would be unidiomatic.

### Rendered Incorrectly

An example of an everyday term is *situaciones consolidadas*, which we documented in EUR-Lex as ‘previous situations’. GT’s word-for-word rendition ‘consolidated situations’ is too difficult to make sense of. Another everyday term is the lexically ambiguous *fórmula*, which appears in the phrase *Específica la Sala que no le da a dicho requisito la consideración de fórmula expresa y solemne....* We managed to document *fórmula* in Alcaraz and Hughes (2005) as ‘wording’, as in ‘The Chamber specifies that it does not give consideration to that requirement in express and solemn wording...’. GT, however, made the mistake of translating it as ‘formula’, which in most other contexts would be appropriate.

*Texto Refundido de la Ley de Sociedades Anónimas* is an official commercial law term we documented in EUR-Lex as ‘Consolidated Version of the Law on Public Limited Companies’. GT, on the other hand, came up with ‘Consolidated Companies Law’, which appears in a good number of sentence examples crawled by Linguee (which can be thought of as a sort of ‘Google of bilingual text’). A potential problem might be that its inclusion of ‘Companies’ and not ‘Public Limited Companies’ is unnecessarily too general a solution, whereas ‘company’ is normally the equivalent of *sociedad* when it is unmodified by an adjective and ‘public limited company’, normally that of *sociedad anónima* in the European context (e.g. in the United States one might prefer ‘joint-stock company’). Nevertheless, as the EUR-Lex database continues to grow it is now possible to find more English translations of the source term that employ ‘Companies’ instead of ‘Public Limited Companies’ (e.g. ‘Amended Text of the Law on Companies’). So perhaps our categorization is nitpicky or our hand-picked solution runs the risk of being clumsy or too long to be practical.