

# **HONR 218L – Language and Mind**

Patterns of failure in machine translation

J. Ozymandias Fallick

## **Honor Pledge**

I pledge on my honor that I have not given or received  
any unauthorized assistance on this assignment.

## **Introduction**

The Internet reaches people all over the globe. Users live in hundreds of countries and speak many languages. Wikipedia has editions in 276 languages.<sup>1</sup> In the global society that the Internet provides, exposure to material in other languages is inevitable, and the need for quick and accessible translation arises. Because it eliminates the need for a human interpreter, returns fast results, and can be accessed on a local computer or over the Internet, machine translation (MT) can fill this niche.

Google provides an MT tool called Google Translate (accessible at <http://translate.google.com/>), which has become extremely popular, receiving 1% of Google's enormous traffic.<sup>2</sup> I can think of several possible causes for Google Translate's popularity. Google is the world's most popular website,<sup>2</sup> and frequent Google users may be more likely to go to Google for translation than to a different website. Google Translate offers 57 languages,<sup>3</sup> making it very versatile, and it integrates with other Google products, like Google Search and the Google Chrome web browser. Perhaps most appealing is that it is free of charge, not even requiring a Google account for use.

## **How Google Translate works**

Unlike some MT tools, which attempt to translate between languages using preprogrammed rules, Google Translate uses statistical MT, which entails detecting patterns in multilingual documents in which text in one language has been translated to another by a human. As a result, it is not necessary for all of a language's rules to be programmed in ahead of time, and the quality of the translation varies with the amount of material available.<sup>3</sup> This method is especially suitable for Google, which already has vast numbers of documents indexed for its primary Search service.

## **When Google Translate falls short**

To examine shortcomings of MT, I used Google Translate to translate phrases between English, my native language, and French and Latin, two languages I have studied, and analyzed what aspects of those phrases caused errors in the translations. The documents used for translation were the English Wikipedia article on the Liberty Bell,<sup>4</sup> the French Wikipedia article on Mohandas Karamchand Gandhi,<sup>5</sup> and the Latin Wikipedia article on the Rosetta Stone.<sup>6</sup> All three were Featured Articles, and were chosen to avoid prevalent technical language. To prevent bias, it is beneficial to translate documents for which

direct translations do not exist on the Internet. Wikipedia Featured Articles are, by virtue of their featured status, sizable, well-written compositions, and, since they are dynamic and mutable documents, they are unlikely to have direct translations published on the Internet. While Featured Articles are likely to have equivalents in other languages, they are generally not direct translations; to ensure that this is the case, I checked each article's equivalent in the other languages to verify that they are in fact not direct translations.

### *English-French-English*

The translation quality between English and French is very good. Since there are an estimated 536.6 million Anglophones and 59.8 million Francophones on the Internet,<sup>7</sup> Google Translate has plenty of material to work with, and Google has heavy incentive to support French.

The following categories of errors occurred in a translation of the Liberty Bell Wikipedia article from English to French to English.

### Equivalents

When two English words or phrases are synonymous, and have the same translation in French, one may be substituted for the other in the roundtrip translation. This doesn't make the translation incorrect; it only deviates from the original in diction.

- (1) the bell was **commissioned** from the London firm → la cloche a été **commandée** à la firme londonienne → the bell was **ordered** from the London firm

In sentence (1), *commissioned* and *ordered* can both be translated into French as *commandé*, so *ordered* was substituted for *commissioned* in the translation.

- (2) Proclaim LIBERTY throughout **all** the land **unto** all the inhabitants **thereof**. → publiez la liberté dans **toute** la terre **pour** tous **ses** habitants. → proclaim liberty throughout the land **to** all **its** inhabitants.

This effect is not restricted to words alone. Phrases can change to equivalents as well. In sentence (2), *all the inhabitants thereof* becomes the equivalent *all its inhabitants*. The *all* in *throughout all* is dropped because the corresponding French phrase *dans toute* can also be translated as *throughout*, and *unto* becomes *to* because both are translated with *pour* in this context.

### Equivalent in French but not English

Again, when two English words or phrases have the same translation in French, one may be substituted for the other in the roundtrip translation. However, if they are not equivalent in English, this can change the meaning of the sentence.

- (3)            was **cast** with the lettering → et a été **coulé** avec le lettrage → was **sank** with the lettering

In sentence (3), *sank* is substituted for *cast* because in French, the verb *couler* is used for both *to sink* and *to cast (metal)*. Since these verbs are not interchangeable in English, the meaning of the sentence is changed.

- (4)            whose **last** names appear on the bell → dont les noms apparaissent sur la cloche → whose names appear on the bell

Sometimes the specificity of the original English can be lost. In sentence (4), *last* is dropped from *last names* because in French, the word *nom* can be used to refer to *name* in general or to *last name* specifically.

### Gender

Since French is a gendered language, but in English, gender corresponds almost exclusively with anatomical gender, translation of pronouns is tricky.

- (5)            **It** acquired **its** distinctive large crack → **Il** a acquis **sa** grande fissure distinctif → **He** acquired **his** distinctive large crack

In sentence (5), *it* and *its* in the original English sentence convey that the object in question is without anatomical gender. However, *il* in French conveys only that the object is of masculine grammatical gender in French, and *sa* conveys nothing about the object's gender. As a result, the translation back to English must choose between *he/his* and *it/its*, because both are possible translations of *il* and *sa*. In this situation, the wrong one is used.

### Change of voice

- (6)            Bells **were rung** to mark the reading → Cloches **sonnaient** à l'occasion de la lecture → Bells **rang** to mark the reading

In sentence (6), the passive *were rung* becomes the active *sonnaient*, which then preserves its voice in becoming *rang*. I'm uncertain as to why the change occurred, but

the same meaning is conveyed, and it's possible that the translator considered the active to be a more common means of expressing the idea in French.

#### Change of tense

Because tenses are used differently in English and French, changes in the form of the tense can occur over the roundtrip translation.

- (7) the bell **was** adopted as a symbol → la cloche **a été** adopté comme symbole  
→ the bell **has been** adopted as a symbol

The English simple past and present perfect, when they both represent past time and perfective aspect, are both translated into French with the present perfect, though they do not have exactly the same meaning. An example of how this changes the translation is in sentence (7), where the simple past *was* is translated into the present perfect *a été*, which becomes the present perfect *has been*. This is similar to the error from equivalents in French mentioned above.

#### Change of word order

Word order is different in English and French, so it is possible for the word order to change over the roundtrip translation.

- (8) a **widespread** story **claims it** cracked → une histoire **répandue prétend**  
fissuré → a story **claiming widespread** cracked

In sentence (8), a word order change occurs. I am unsure as to the cause, but it is possible that the *it* was erroneously dropped in the translation to French, where *widespread* moved to after *story*, since adjectives generally follow verbs in French. Then, because of the lack of a subject and auxiliary for the participle *fissuré*, the translator misinterpreted *répandue* to be modifying *fissure* instead of *histoire*.

#### Others

There were some idiosyncratic errors that I could neither classify nor explain.

- (9) to summon **lawmakers to** legislative sessions **and** to alert citizens **to** public meetings and proclamations → pour appeler le **législateur** à sessions législatives **et** à sensibiliser les citoyens **à** des réunions publiques et les proclamations → to summon **the legislature and** legislative sessions to educate citizens **in** public meetings and proclamations

In sentence (9), the original English was properly translated to French; the translation from French, however, dropped an *and* and changed *to* to *and* and *to* to *in* in English.

- (10) believe it was one of **the bells rung** → pensent qu'elle a été l'un des **sonner les cloches** → believe it was one of **ringing the bells**

In sentence (10), the translation to French changed the form (from participle to infinitive) and position (from after to before *the bells*) of the verb *to ring*. The translation back to English was true to the French, but because the French was wrong, the return to English was also wrong.

#### *French-English-French*

In the translation of the French Wikipedia article on Mohandas Karamchand Gandhi from French to English to French, instances of substitution of equivalents (both the variety that changes the meaning and the variety that does not) and change of word order occur.

Other errors, which did not occur above, appear below.

#### Failure to translate

There were some instances in which a word failed entirely to translate.

- (11) est un **dirigeant** politique, important **guide** spiritual → is a political **leader**, an important spiritual **leader** → est un **leader** politique, un **leader** spirituel important

In sentence (11), the words *dirigeant* and *guide* translate to English as *leader*, but they fail to translate back to French, remaining *leader* in the final sentence.

#### Change of tense

Problems with translation of tense occurred, both as a result of different uses of tense in English and French, and as a result of how tense works in French.

- (12) Il **faisait** ses propres vêtements → He **made** his own clothes → Il **a fait** ses propres vêtements

This error is similar to that in sentence (7). The present perfect, simple past, and imperfect in French (here, *faisait* is imperfect) are all translated to the English simple past (*made*), which is translated back into French as the present perfect (*a fait*).

- (13) Gandhi **développa** une méthode → Gandhi **developed** a method → Gandhi **a développé** une méthode

In French, the simple past and the present perfect are both used to represent past time and perfective aspect. Both are translated into English as the simple past, so the translator cannot make a distinction translates the verb into French as the more common present perfect.

#### Gender agreement

In French, the modifiers of nouns must agree with the gender of the nouns they modify. This is not always preserved in the translation.

- (14) Cette date a été **déclarée** → This date was **declared** → Cette date a été **déclaré**

Past particles following “to be” used as an auxiliary verb must agree with the subject of the sentence in French. In sentence (14), the final e is dropped from *déclarée* on the return translation, removing the agreement. I believe that this is because the sentence was present perfect and passive, so the phrase *été declare* (*été* being the past participle of *être*, “to be”) had *a*, the conjugated form of “to have,” as its auxiliary, which may have confused the translator. In other instances, when the agreement was due to adjective use or present perfect using “to have” as the auxiliary, the translator preserved the agreement.

#### *English-Latin-English*

Unlike French, Latin is a language marked as “alpha” on Google Translate, indicating that it was added recently, and that the translation quality is not expected to be good. In addition, while French has a vast and vibrant community of speakers on the Internet, Latin is a dead language, with a very limited population of speakers. Most of the material that Google Translate can use for Latin is Roman literature and old scientific treatises. While Latin documents on the Internet are more likely to exist alongside an English translation than French documents, there are far fewer of them, so Google Translate has much less to work with. As a result, the translation is quite poor.

Because the translation is so fraught with errors, I will break from the format of the previous section and give a listing of the common errors, then analyze example sentences, as each sentence contains several errors. The most problematic error is a poor

Latin vocabulary. Latin nouns are inflected for case, which is vital to the meaning of a sentence. When the translator can find no translation into Latin for an English noun, it is unable to decline the noun, leaving the sentence difficult to translate into English correctly. This also causes errors in word order, as Latin has a far more flexible word order than English, since the function of words in Latin sentences is given by inflection instead of position. Other issues arise from the lack of articles in Latin, which causes the translator to have to guess where to place articles, gender of Latin nouns, which causes errors similar to those in French, and the greater versatility of words in Latin, which leads to frequent equivalent-caused errors.

The following sentences are from the translation of the Liberty Bell Wikipedia article from English to Latin to English.

- (15) The Liberty Bell is an iconic symbol of American Independence, located in Philadelphia, Pennsylvania. → *Libertas est ICONICUS Bell symbol of American Independence, site in Philadelphia, Pennsylvania.* → Freedom is the Bell ICONIC symbol of American Independence, located in the Philadelphia, Pennsylvania.

In sentence (15), the word *Bell* and the phrase *symbol of American Independence* fail to translate to Latin. As a result, *Bell* does not decline, and the translator cannot tell that *Libertas* is supposed to modify *Bell*. On the translation back to English, *Libertas* is treated as its own noun, yielding *Freedom is the Bell* instead of *Freedom Bell is the*, which would preserve the original function of the nouns. In addition, the initial *the* is lost and an erroneous *the* is inserted twice in the roundtrip translation, due to Latin's lack of articles.

- (16) In its early years, the Liberty Bell was used to summon lawmakers to legislative sessions and to alert citizens to public meetings and proclamations. → *Suo primis annis libertatem solebat Bell vocant ad legislatores et legislativa alert sessions cives edictum contionum.* → By his own during the first years of war call used to the freedom of legislators and to the legislative sessions alert the citizens of the edict of its assemblies.

Sentence (16)'s translation opens with a declension error. *In its* was translated to *suo* instead of *suis*, thus failing to agree in number with *primis annis* ("in first years") and causing the translation back to English to treat *its* (which becomes *his* due to a gender



error) and *first years* as two separate phrases. *Bell* fails to translate again, and is dropped entirely from the return to English. *To summon* is translated into the third person plural instead of the infinitive in Latin, then somehow becomes the subject of its phrase, since no noun in the nominative (subjective) case is available: *Bell* is undeclined, and *Liberty* is translated into the accusative (direct objective) case instead of the correct nominative case. Since *Liberty* is translated into the accusative case, the translation back to English treats it as the object of the phrase, and *lawmakers*, though it is translated into the correct case (accusative) in Latin, becomes possessive in the English translation. *Sessions* is untranslated, and both it and *legislative* are placed on the wrong side of the conjunction in the Latin translation, and so *legislative sessions* is in the wrong clause in the English translation. *Alert* is also untranslated, but the resultant English translation preserves the meaning, save for a missing conjunction between *public meetings* and *proclamations* and loss of the plural on the translation from Latin to English.

#### *Latin-English-Latin*

Translation from Latin to English to Latin appears to be better at preserving the original form; however, the English stage of the translation continues to be poor. The following sentences are from the translation of the Latin Wikipedia article on the Rosetta Stone from Latin to English to Latin.

- (17) Tabula Rosettana[1] (etiam titulo OGIS 90 recognita) est stela decreto de rebus sacris in Aegypto anno 196 a.C.n. lato inscripta. → Table of the Rosettana [1] (even has reviewed the title of OGIS 90) is a decree of the COLUMN concerning sacred things in Egypt, in the year 196 BC with a broad Inscribed. → Tabula Rosettana [I] (etiam recognita OGIS nomen XC) est columna decretum de rebus sacris in Aegypto anno CXCVI lato BC Inscriptae.

The roundtrip translation of sentence (17) is far more accurate than (15) or (16). The only errors are three losses of case: *titulo* (“title”) goes from ablative to nominative (and is replaced by the equivalent *nomen*), *decreto* (“decree”) goes from ablative to nominative, and *inscripta* (“inscribed”) goes from nominative singular to genitive singular, dative singular, or nominative plural (the forms are identical).

- (18) Inventio stelae, linguis duabus scripturisque tribus inscriptae, eruditus Instituti Aegypti statim nuntiata est → The Finding of and stars, the two scripturisque

the tongues of the tribes of the register, the learned of the Institute of Egypt is told at once → *Inventio sidera duo linguas scripturisq̄ue tribus inscriptae, eruditis Instituti Aegypti statim nuntiata est*

The first error in sentence (18) appears to be an equivalency error and a case error; *stelae* is translated as *stars* instead of *columns*, which results in the use of a different Latin word meaning only “stars” (*sidera*) in the final Latin translation. In addition, *stelae* is interpreted as nominative plural instead of the proper genitive singular, which has an identical form for the word *stela*. *Sidera*, however, would be *sideris* in the genitive singular, so the Latin translation does not preserve the case. The only other error in the sentence is the change of case of *linguis* (“in the languages”) from ablative to accusative and of *duabus* (“two”) from feminine ablative to masculine nominative, neuter nominative, or neuter accusative (the forms are identical). This is reflected in the English translation, which lacks the “in” implied by the use of the ablative case.

### **Conclusion**

Because Google Translate uses statistical machine translation, a large determining factor in the quality of the translation is the volume of documents available in which to find patterns. Since English-French has many documents on the Internet to draw from, the quality is generally very good, with occasional errors that can generally be isolated and explained. English-Latin, however, has a far smaller corpus to use, and as a result the translation is far poorer, and errors are pervasive. This is also a property of Latin’s heavy inflection, which may have to do with why Latin-English-Latin is so much better than English-Latin-English: because a single form in Latin often has a meaning based on context and can be translated several ways into English, the intermediate English stage of the translation may be translated the wrong way, but on the return to Latin the incorrect translation takes the same form as the correct translation would. The quality of Google Translate for languages as popular as French and English is fantastic for machine translation, and it can be a useful and valuable tool for the web surfer who comes across a page in the other language, and needs a quick translation. An Anglophone encountering Latin is out of luck, since Google Translate from Latin to English does not preserve meaning well. A Latin-speaking web surfer needing to read an English page may be more fortunate, but Latin-speaking web surfers are in short supply.

---

<sup>1</sup> Wikimedia Meta-Wiki. 2010. *List of Wikipedias*. Retrieved from

[http://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](http://meta.wikimedia.org/wiki/List_of_Wikipedias)

<sup>2</sup> Alexa. 2010. *Google.com Site Info*. Retrieved from <http://www.alexa.com/siteinfo/google.com>

<sup>3</sup> Google. 2010. *Inside Google Translate*. Retrieved from [http://translate.google.com/about/intl/en\\_ALL/](http://translate.google.com/about/intl/en_ALL/)

<sup>4</sup> Wikipedia. 2010. *Liberty Bell*. Retrieved from

[http://en.wikipedia.org/w/index.php?title=Liberty\\_Bell&oldid=401147651](http://en.wikipedia.org/w/index.php?title=Liberty_Bell&oldid=401147651) (permanent link)

<sup>5</sup> Wikipédia. 2010. *Mohandas Karamchand Gandhi*. Retrieved from

[http://fr.wikipedia.org/w/index.php?title=Mohandas\\_Karamchand\\_Gandhi&oldid=59884063](http://fr.wikipedia.org/w/index.php?title=Mohandas_Karamchand_Gandhi&oldid=59884063) (permanent link)

<sup>6</sup> Vicipædia. 2010. *Tabula Rosettana*. Retrieved from

[http://la.wikipedia.org/w/index.php?title=Tabula\\_Rosettana&oldid=1391337](http://la.wikipedia.org/w/index.php?title=Tabula_Rosettana&oldid=1391337) (permanent link)

<sup>7</sup> Internet World Stats. 2010. *Internet World Users by Language*.

<http://www.internetworldstats.com/stats7.htm>