

CS 224n - PA 1 - Google Translate Analysis

Jonathan NeCamp, Andrew Giel

October 10, 2014

Translations

1)

F: le parlement n' a pas ratifié l' amendement pour la libération de tymosenko

Our System: parliament has not ratified the amendment for the liberation of tymosenko

Google Translate: parliament has not ratified the amendment to the release of Tymosenko

2)

F: le parlement ukrainien a refusé , dans la cadre d' un amendement au droit pénal , le projet d' annulation du paragraphe relatif à l' inculpation de julia tymosenko , la chef de l' opposition

O: the ukrainian parliament has refused , in the context of an amendment to the criminal law , the project cancellation paragraph establishing charges of julia tymosenko , the head of the opposition

GT: Ukrainian parliament refused, as part of an amendment to the criminal law, the proposed cancellation of the paragraph relating to the charge julia Tymosenko, the leader of the opposition

3)

F: les députés ont refusé en deuxième lecture le projet de modification visant la réduction des peines pour délits économiques , qui aurait pu ouvrir les portes de la liberté pour l' ex-première ministre actuellement emprisonnée

O: members have refused to second reading of the proposed amendment reduction of penalties for economic crimes , which could open the doors of freedom for the former minister currently in prison

GT: MEPs rejected on second reading the proposed amendment to reduce sentences for economic crimes, which could open the doors of freedom for the former Prime Minister currently imprisoned

4)

F: tymosenko a été condamnée en octobre à 7 ans de prison pour avoir conclu un accord à priori dés avantageux avec la russie pour l' achat de gaz naturel

O: in october tymosenko was sentenced to seven years in prison for reaching an agreement in essence disadvantageous with russia for the purchase of natural gas

GT: tymosenko was sentenced in October to 7 years in prison for an agreement in principle with russia dice advantageous for the purchase of natural gas

5)

F: le jugement n' est pas définitif et le tribunal doit statuer sur l' appel de la condamnée en décembre

O: the ruling is not final and the tribunal must decide on the appeal of the sentenced in december

GT: the judgment is not final and the court must rule on the appeal of the convicted in December

6)

F: tymosenko qualifie le jugement de vengeance politique du régime et a provoqué un processus de soupçons de partialité du tribunal également à l'ouest

O: tymosenko described the judgement of political revenge and caused a process of suspicion of the court also in the west

GT: Tymosenko qualifies the judgment of political revenge of the regime and led a process of suspicions of bias court also west

7)

F: le projet d'annuler le paragraphe 365 du droit pénal, sur la base duquel l'ex-première ministre a été condamnée, a été signé par 147 députés

O: the project to cancel the article 365 criminal law, on the basis of which the minister was sentenced, was signed by 147 deputies

GT: the project to cancel the paragraph 365 of the criminal law on the basis of which the former first minister was sentenced, was signed by 147 members

8)

F: il aurait fallu 226 voix pour l'approuver

O: it would have taken 226 votes to approve it

GT: would have needed 226 votes for approval

9)

F: victoire libyenne

O: victory libyan

GT: Libyan victory

10)

F: la libération ou rébellion libyenne a déjà ses vaincus

O: the release or rebellion libyan has already overcome

GT: Libyan liberation or rebellion already has his vanquished

11)

F: muammar kaddafi est enterré dans un endroit inconnu dans le désert et sans lui la guerre est terminée

O: muammar kaddafi is buried in an unknown location in the desert and without him the war is over

GT: muammar Gaddafi was buried at an undisclosed location in the desert and without him the war is over

12)

F: il reste à déterminer les vainqueurs

O: it remains to determine the winners

GT: it remains to determine the winners

13)

F: comme il est de coutume dans la région, les vainqueurs aux élections sont les islamistes, mais s'agit-il des modérés ou des radicaux

O: as is customary in the region, the winners in elections are the islamists, but is this of moderates or radicals

GT: as is customary in the region, the winners in the elections are Islamists, but is a question -there moderates or radicals

14)

F: le conseil national provisoire a décrété le droit habituel de la charia et nous savons déjà de quoi il s'agit

O: the interim national council has the right of sharia and we already know what it is
GT: the provisional national council decreed the usual sharia law and we already know what it is a question

15)

F: la libye devient un pays sans criminalité car pour un vol on coupe la main

O: libya becomes a country with no crime because for a flight is cut the hand

GT: Libya is a country without crime because a flight is cut by hand

16)

F: par exemple dans le désert vous vous retrouvez dans un tempête de sable qui vous diminue de façon sensible la visibilité

O: for example in the desert you find yourself in a sandstorm you declines significantly the visibility

GT: for example in the desert you find yourself in a sand storm that reduces your visibility significantly

17)

F: les épreuves sont en plus de temps en temps agrémentées d' un élément atypique

O: the prints are more time to time , is an element

GT: tests are in addition to occasionally embellished with a unique element

18)

F: mais la conduite à travers le désert , dans des paysages d' automne ou bien dans des montagnes hautement enneigées est très agréable visuellement

O: but the actions across the desert , in autumn or in mountains highly snow-covered is very pleasant visually

GT: but driving through the desert landscape in autumn or in high mountain snow is very visually pleasing

19)

F: enfin ce n' est pas le bout du monde

O: lastly , this is not the end of the world

GT: Finally, this is not the end of the world

20)

F: certains environnements sont de surcroît assez jolis même si le niveau technique graphique est du même type que les séries précédentes

O: certain environments are worries enough nice even if the level graphic technique is the same type as the previous rounds

GT: some environments are further nice enough although the technical level graph is the same type as the previous series

21)

F: le joueur n' a pas le sentiment un seul instant que le jeu lui propose quelque chose d' identique

O: the player does not feel that the proposes something of identical

GT: the player does not feel for a moment that the game offers him something identical

22)

F: ce qui en résulte que chaque course se fait sur un autre itinéraire et les étapes prévues dans des environnements différents qui se distinguent assez bien les uns des autres et sont intéressants

O: this fact that each race on a different route and the steps in different environments who stand out quite well of each other and are interesting

GT: the result is that each race is on another route and the planned steps in different environments that stand well enough from each other and are interesting

23)

F: l' itinéraire est divisé en dix étapes et ensuite en plus de 50 courses

O: the route is divided into ten steps and then in over 50 races

GT: the route is divided into ten steps and then over 50 races

24)

F: l' idée a des éléments intéressants

O: the idea of interesting points

GT: the idea has interesting elements

25)

F: vous vous asseyez derrière le volant sur la côte ouest à san francisco et votre mission est d' arriver le premier à new york

O: you sit down behind the wheel on the west coast in san francisco and your mission is to get to the first to new york

GT: you sit behind the wheel on the west coast in San Francisco and your mission is to be the first at new york

26)

F: vous êtes pris dans l' aventure dont le héros principal jack a des problèmes caractériels et pour les régler se met à participer à des courses illégales à travers l' Amérique

O: you are caught in the adventure the lead jack problems personality and to pay will begin to participate in races illegal across america

GT: you are caught in an adventure whose hero Jack has characterological problems and resolve begins to participate in illegal races across america

27)

F: l' aventure commence très bien

O: the adventure begins very well

GT: the adventure begins well

28)

F: ce ne sont pas vraiment des erreurs mais plutôt l' expression d' une fainéantise incompréhensible de la part des développeurs

O: they are not really mistakes , but the expression of an incomprehensible laziness on the part of the developers

GT: these are not really mistakes but rather the expression of an incomprehensible laziness on the part of developers

29)

F: ce sont potentiellement des jeux de course d' arcades mais l' ensemble tire vers le bas à cause de quelques fautes inutiles qui existent déjà depuis plusieurs années

O: they are potentially racing games arcade but all drags down a few grammar unnecessary that already exist for several years

GT: are potentially racing games arcades but all pulls down because of some unnecessary mistakes that have existed for many years

30)

F: comme cela se fait déjà dans la série nfs , le résultat est complètement autre que ce que l' on attendait

O: as has already been done in the series nfs , the outcome is completely more than what

GT: as is done already in the NFS series, the result is completely different than what was expected

Analysis

While analyzing these translations, one notices a variety of errors on behalf of both our system and that of Google Translate, yet it is clear that Google does a much better job at giving more nuanced and articulate translations. The errors observed in this sample of translations can be broken down into three categories: Ordering errors where words or phrases are misordered, Translational errors where one word or phrase is incorrectly translated into an incorrect word or phrase, and Fluency errors where the translation as a whole is lacking the flow of a proficient speaker of the language.

In terms of ordering errors, Google outperforms our system by a long shot. The most salient example of this is (9), but more minor examples (26) are throughout the sample corpus. This can either be a result of Google's use of a POS tagger or simply more data to train their phrase based system.

Translational errors seem to be present in both systems, yet it is hard to say with high certainty without being fluent in French (which neither of us are). This being said, it does seem that Google once again is superior to our system. Often our system and Google's express the same statement but simply use slightly different words in the process, with Google's choice of words seeming very subtly better. A few examples of this include (1), (5), and (11). There are other examples such as (17) and (29) as well that exhibit bad French to English mappings. This is most likely due to massive amounts of data Google is able to train its model on, as it has a much larger vocabulary to choose from when making translations.

At this point in the analysis and comparison of the two systems it seems that Google is ahead, but not by much. Although the last category of errors, Fluency errors, is where Google really pulls ahead. Fluency errors are really just holistic judgements of the translations in the eyes of native speakers of English (which we both are). It is apparent through this lens that Google has a better system for translation. Nearly every example in this small corpus is a tribute to the fluency of Google's translations (especially compared to ours), yet (2) and (16) especially emphasize this difference. This, once again, is most likely due to the amount of data at Google's disposal as they can have a much more accurate English language model to leverage in their overall system.

In conclusion, our system is lacking when compared to that of Google's in its ability to exhibit knowledge of nuance and fluency in language, yet still gives translations that serve to deliver the same overarching message as Google Translate.