

Machine Translation Homework 1

Group members: Chen, Tianyi; Tan, Kanran (Ordered By First Letter Of Last Name)
tchen59@jhu.edu, ktan7@jhu.edu

#1

1. Source sentence: 对于英国夫妻捷德和利亚姆来说，大概再也收不到比刚出生的女儿 Poppy 送给他们的更好的情人节礼物——一个印在前额的心型胎记。

2. Machine translation: For the Czech Republic, Germany and UK couples Liam, it probably could not receive more than just baby daughter Poppy gave them better valentine, printed on a heart-shaped birthmark on the forehead.

3. Correction: For English couple Jade and Liam, they probably could not receive a better Valentines' gift than a newborn daughter Poppy—a heart-shaped birthmark on her forehead.

4. Assessment:

1) Noun translation is wrong: “捷德” should be combined and translated as a person name of the couple, Jade, instead of “Czech Republic, Germany”. In the Machine translation, “捷德” is separated in two words, which “捷” represents “Czech Republic” and “德” represents Germany.

2) Wrong Subject: “they” should be the subject instead of “it” at the beginning of the sentence “it probably could not receive more than just baby daughter Poppy gave them better valentine”.

3) Two verbs in one sentence: For sentence “probably could not *receive* more than just baby daughter Poppy *gave* them better valentine”, “gave” is mistakenly added into sentence.

4) Missing noun: “gift” is missing after “better valentine”

5) Wrong format of adverbial: “printed on a heart-shaped birthmark on the forehead”. “Printed” is a verb to further describe “birthmark” along with “on the forehead” as adverbial but instead directly functioning on “birthmark”. It would make more sense if the sentence was rearranged as “a heart-shaped birthmark printed on the forehead”. Nevertheless, the word “printed” could have been better translated into another word.

#2

1. Source sentence: 据英国《每日邮报》2月11日报道，20岁的捷德和21岁的Liam是一对来自英国林肯郡的年轻夫妻。

2. Machine translation: According to the British “Daily Mail” reported on February 11, Czech Republic, Germany and 21-year-old Liam 20-year-old is a pair of young couples from the United Kingdom, Lincolnshire.

3. Correction: According to the British “Daily Mail” reported on February 11, 20 year-old Jade and 21 year-old Liam are young couple from the United Kingdom, Lincolnshire.

4. Assessment:

1) Noun translation is wrong: “捷德” should be combined and translated as a person name of the couple, jade, instead of “Czech Republic, Germany”, which are countries. In the Machine translation, “捷德” is separated in two words, which “捷” represents “Czech Republic” and “德” represents Germany.

2) Wrong position of adverbial: “20-year-old” should be placed right after “...February 11,” but since “Jade” was unable to be translated as a name, the translation system might not be able to figure out the right place to put this adverbial.

3) Redundant word or phrase usage: “a pair of young couples”, here we already have couple, so it’s confusing to add “a pair of” before the “couple”. But this happened because there is something similar to “a pair of” in the source sentence, due to the Chinese syntactic grammar.

#3

1. Source sentence: 去年 11 月份二人迎来了第一个孩子，女儿 Poppy。

2. Machine translation: Last November, the two welcomed their first child, daughter Poppy.

3. Correction: Last November, the couple welcomed their first child, daughter Poppy.

4. Assessment:

1) Word sense error for noun: “The two” is completely word translated into English but it should be translated as “the couple”.

#4

1. Source sentence: 与众不同的是，小家伙额前自带了一块心型的胎记。

2. Machine translation: Unusually, the amount before the little guy comes with a heart-shaped birthmark.

3. Correction: Surprisingly, the little baby came with a heart-shaped birthmark on her forehead.

4. Assessment:

1) Tenses error: “comes” should be “came” since this event was happened in the past.

2) Improprate Adverb. “Unusually” is always used before an adjective instead of using at the very beginning of a sentence. Here “surprisingly” might be better by given context.

3) Word sense error for noun: “the amount before” should be translated as “forehead”.

#5

1. Source sentence: 捷德是一位全职妈妈。

2. Machine translation: G & D is a full-time mom.

3. Correction: Jade is a full-time mom.

4. Assessment:

Noun translation is wrong: “捷德” should be translated as a person name, Jade, instead of “G & D”.

#6

1. Source sentence: 她说女儿刚出生几小时后二人就发现了 Poppy 额前有一小块红色印记。

2. Machine Translation: She said that after her daughter was born a few hours before the two found a small amount of Poppy red mark.

3. Correction: She said that a few hours later after her daughter was born the couple found a small red mark on Poppy’s forehead.

4. Assessment:

1) Reordering error: In the sentence of “after her daughter was born a few hours before”, the “a few hours” should be placed before “her daughter was born”. This is due to the syntactic grammar difference between Chinese and English that in Chinese it the “few hours later” or “few days ago”, etc., are always added after a specific time; but in English the situation is somehow reversed.

2) Wrong preposition: “a few hours before” should be “a few hours later”.

3) Word sense error for noun: “The two” is completely word translated in English but it should be better translated into “the couple”.

4) Word sense error for noun and Wrong position of adverbial: “amount of ” should be translated as “forehead” and it should be place after “red mark” as its adverbial

#7

1. Source sentence: 医生称那是生产时留下的压痕，几小时后就会自然消失。

2. Machine Translation: Doctors said that the production of the indentation left, after a few hours it will disappear naturally.

3. Correction: Doctors said that was the indentation left in parturition, and after a few hours it would naturally disappear.

4. Assessment:

1) Tenses error: “will” should be “would”.

2) Missing predicate: In sentence of “Doctors said that the production of the indentation left”, there is no predicate. If add “be” after “that”, it would meet the requirement of a complete sentence.

3) Translating to improper word: “the production” should be better translated. “Parturition” might be a good choice.

4) Reordering error: In the sentence of “Doctors said that the production of the indentation left”, “the production (parturition)” should be placed after “left”, combining with “left” to give more information of “indentation” and playing role of adverbial instead of a subject.

5) Reordering improvement: “it will disappear naturally” would be much better if translating to “it would naturally disappear”.

6) Run-on sentence: “Doctors said that the production of the indentation left, after a few hours it will disappear naturally.” This sentence indicates it has two parts which is separated by the comma. However, it should use “and” to connect the two parts otherwise it becomes a run-on sentence.

#8

1. Source sentence: 然而两天后夫妻发现红印不仅没有消失，反而逐渐变成了心型。

2. Machine translation: However, two days later the couple found red marks not only not disappeared, but has gradually become a heart-shaped

3. Correction: However, two days later, the couple found out that the red mark did not disappear, instead gradually becoming a heart-shaped.

4. Assessment:

1) Tenses mismatching: Two verbs “found” and “has gradually become” are not matching with each other’s tenses.

2) Repeated negation usage: “not only not...but” can be better translated by using “instead”.

#9

1. Source sentence: 捷德担心孩子的健康便带着女儿咨询了医生。

2. Machine translation: Giesecke & Devrient worried about your child’s health so she took her daughter to the doctor consultation.

3. Correction: Jade worried about her child's health so she took her daughter to consult doctors.

4. Assessment:

1) Noun translation is wrong: “捷德” was translated to “Giesecke & Devrient”, which should be a person name “Jade”

2) Wrong Possessive Adjectives: “your child” should be translated to “her child”. The wrong possessive adjective may be caused by the ambiguous noun “捷德” that the translation system did not know it's a name of a person and so further misunderstand the relationship of “捷德” and the child

3) Using verb rather than noun: Though the sentence “she took her daughter to the doctor consultation” has no grammar mistakes or anything very confusing, but it can be better translated to “she took her daughter to consult doctors” by using verb instead of noun.

#10

1. Source sentence: 而医生再次确认“红心”只是一块普通的胎记，并无大碍。

2. Machine translation: The doctors confirmed that “Red ” is just an ordinary birthmark again, it does not matter.

3. Correction: The doctors confirmed that the “Red heart” was just an ordinary birthmark again, and it did not matter.

4. Assessment:

1) Tenses errors: “is” should be “was” and “it does not matter” should be “it did not matter”

2) Missing article: “the” should be added before “Red ..” which will clearly indicate the specific “Red ...” otherwise it will cause grammar mistake of ambiguity.

3) Run-on sentence: “The doctors confirmed that “Red ” is just an ordinary birthmark again, it does not matter.” This sentence indicates it has two parts, which are separated by the comma. However, it should use “and” to connect the two parts otherwise it becomes a run-on sentence.

#11

1. Source sentence: 看见心型痕迹的时候我们既吃惊又害怕，因为不知道是不是由于健康问题造成的。

2. Machine Translation: See the heart-shaped mark when we both surprised and afraid, because they do not know whether it was due to health problems caused.

3. Correction: Seeing the heart-shaped mark, we were both surprised and afraid, because we did not know whether it was due to health problems.

4. Assesment:

- 1) Tense errors: “do not know” should be “did not know”
- 2) Wrong Subject(s):
 - a) “we” should be the correct subject, while there is no subject in “See the heart-shaped mark when we both surprised and afraid”.
 - b) “we” should be the correct subject, while “they” is the subject in “because they do not know whether it was due to health problems caused”.
- 3) Predicate is wrong:
 - a) “be” should be the correct predicate, while “see” is the predicate in “See the heart-shaped mark when we both surprised and afraid.” The correct sentence should be “Seeing the heart-shaped mark, we are both surprised and afraid”.
- 4) Redundant component:
 - a) “caused” is redundant in “due to health problems caused”. “due to health problems” can already express “由于健康问题造成的” sufficiently.

#12

1. Source sentence: 后来更是意外发现女儿原来是去年情人节时怀上的。
2. Machine Translation: Later it is discovered by accident when her daughter turned out to be pregnant on Valentine’s Day last year
3. Correction: Later it was discovered by accident that she was pregnant on Valentine’s Day last year, which turned out to be her daughter.
4. Assesment:
 - 1) Tenses: “Later it is” should be “Later it was”.
 - 2) Reordering error: “when her daughter turned out to be pregnant on Valentine’s Day last year” means her daughter was pregnant on Valentine’s Day last year, but actually the person who was pregnant is she(the mother) who owns her daughter. So this part should, instead, be translated to “she was pregnant on Valentine’s Day last year, which turned out to be her daughter”.

#13

1. Source sentence: 这真是一个太可爱的巧合了!
2. Machine Translation: This is really a very lovely coincidence.
3. Correction: This is really a very lovely coincidence.
- 4: Assesment: The translated sentence is totally correct!

#14

1. Source sentence: 对于我们来说, 这个巧合无疑使情人节有了另外一层特殊的意义。

2. Machine Translation: For us, this is undoubtedly a coincidence Valentine had another layer of special meaning.

3. Correction: For us, this coincidence undoubtedly let Valentine have another layer of special meaning.

4. Assessment:

1) Subject is wrong:

a) "coincidence" should be the sentence's subject, not "this"

2) Run-on sentence: Without correcting the original sentence, it was a run-on sentence that "this is undoubtedly a coincidence" and "Valentine had another layer of special meaning" are in the one sentence. So by making "coincidence" as the subject and adding "let" as the predicate, we are able to let this sentence meet the requirements of a complete sentence.

#15

1. Source sentence: 女儿就是我们的小丘比特。

2. Machine Translation: Our daughter is a little Cupid.

3. Correction: Our daughter is a little Cupid.

4. Assessment: The translated sentence is totally correct!

#16

1. Source sentence: 目前已满 12 周的 Poppy 在开心、生气或者难过时，额前的心型胎记都变得更红。

2. Machine Translation: Poppy is currently full 12 weeks in the happy, angry or sad, the heart-shaped birthmark on the forehead have become more red.

3. Correction: Nowadays, when 12-weeks-old Poppy is happy, angry or sad, the heart-shaped birthmark on her forehead becomes redder.

4. Assessment:

1) Reordering errors: "Poppy is currently full 12 weeks in the happy, angry or sad" does not make sense although we could predict its meaning by given words. As our correction, we use "Nowadays, when 12-weeks-old Poppy is happy, angry or sad"

2) Using possessive determiners instead of article: "the heart-shaped birthmark on the forehead" should be "the heart-shaped birthmark on her forehead", where "her" means Poppy here.

3) Tense error and singular or plural: “have become more red” should just be “becomes redder” since we are talking about general situation not an action lasting in a period of time or lasting to the present.

4) Subject is wrong: “the heart-shaped birthmark” should be the subject, not “Poppy”.

5) Doesn’t follow the Rule of one-syllable adjective: “more red” should be “redder”.

#17

1.Source sentence: 一开始我们担心女儿额前的胎记会使她遭到嘲笑, 甚至影响她恋爱交友.

2.Machine Translation: We started to worry about her forehead birthmark she would be ridiculed, and even affect her love friends.

3.Correction: At the beginning, we worried about her forehead birthmark which she would be ridiculed, and which would even affect her having relationship and making friends.

4.Assesment:

1) Word sense error: “一开始” means “at the beginning” but in the machine translation, it turns out using “started” to include the sense of “at the beginning”, while “started” makes no sense.

2) Run-on sentence: “We started to worry about her forehead birthmark she would be ridiculed”, two complete sentences in one sentence, which is not allowed. Since “she would be ridiculed” because of the birthmark, so that the later sentence should be turned into a relative clause to the birthmark.

3) Word sense error: “恋爱交友” means “having relationship and making friends” instead of “her love friends”, and “her love friends” makes no sense here.

#18

1.Source sentence: 但是现在我们意识到, 正是这个心型胎记使她成为了世界上独一无二的人。

2.Machine Translation: But now we realize that it is this heart-shaped birthmark made her unique in the world people.

3.Correction: But now we realize that it is this heart-shaped birthmark that makes her the unique people in the world.

4.Assesment:

1) Run-on sentence: “it is this heart-shaped birthmark made her unique...” here we have two verbs in one sentence. One is “is”, where “it” is its subject and another verb is “made”, where “birthmark” is its subject. Therefore, it’s easy to think about make the second one as a relative clause to birthmark. So we correct it as “it is this heart-shaped birthmark that makes her the unique...”

2) Reordering errors: “made her unique in the world people” should be obviously “makes her the unique people in the world”.

#19

1.Source sentence: 对于我们来说她是最美的。

2.Machine Translation: For us, she is the most beautiful.

3.Correction: For us, she is the most beautiful girl.

4.Assesment:

1) Missing object: Adding “girl” after “the most beautiful” to complete the sentence.

#20

1.Source sentence: 医生称胎记在 4、5 岁左右或者更大的时候将自然消失。

2.Machine Translation: Doctors said the birthmark around 4,5 years old or more when disappear naturally.

3.Correction: Doctors said the birthmark would disappear naturally around 4,5 years old or older.

4.Assesment:

1) No predicate in the clause: For “the birthmark around 4,5 years old or more when disappear naturally”, disappear is the predicate, which makes the sentence looks like “the birthmark disappear naturally around 4,5 years old or more”.

2) Reordering error:

“Around 4,5 years old or older” should be the adverbial, and we should place it behind the predicate. So we change it into “the birthmark disappear naturally around 4,5 years old or older”.

#21

1.Source sentence: 夫妻二人则表示，他们甚至希望这个特别的心型永远不要消失。

2.Machine Translation: The couple said that they even hope that this particular type of heart never disappear.

3.Correction: The couple said that they even hope that this particular type of heart never disappear.

4.Assesment: The translated sentence is totally correct!

Summary: For the Chinese-to-English Machine Translation System(Google translator), it is able to translate and output the key words of each Chinese sentence including noun, verb, adj, etc. So generally, people can easily interpret the meaning from the translation, even if there exist some errors. But after analyzing and comparing the machine translation and correct translation, we can still see some primary quality problems of this system. The first primary issue is translating names. Before translating a Chinese name to an English name, a very critical step might be segmentation that is knowing which portion of a Chinese sentence is a name. In the article that we used, “捷德” is a very common name but the system keeps making mistakes and interpreting it as Czech Republic and Germany, where, in Chinese, Czech Republic is “捷克” and Germany is “德国”. Therefore, it’s easy to see that when the system got the input of “捷德”, it always interpret this word separately by its characters, which may lead it to Czech Republic and Germany since the first Chinese character of these two countries combining as “捷德”. Due to this case, we can probably foresee that in many other cases when the translating system gets an input of a Chinese name, it might interpret the name to something totally irrelative by just looking at its characters. Further, this issue is not just about name translation, it can affect other components’ translation. For example, if a system can’t recognize a name, which is the subject of a sentence, it would be very hard for the system to match up other components of the sentence, such as verb, adverbial and relative clause, and reordering the translated English words, since the system can’t find the correct subject of this sentence which is vital. The second primary issue is word sense errors for noun. When analyzing the system, we found two categories for this problem: one is that the system has no prior information/data for this word, and another is that although the system has these words, when they are combined together, it becomes a new word with completely different meaning that the system is unable to translate. For examples, for the first category, we have “额” in our sentences and the system keeps translating into some wrong words; for the second one, “二人” which means “two” and “person” separately for each character but while combine them together and put into the context of this article, they have a new meaning--“couple”. Because of the above, the system has many word sense errors for noun. The third primary problem is reordering error. This error always happened due to the difference of syntactic grammar between Chinese and English. For example, in English we would say something like “one hour later(before) when something happened” but in Chinese the order is reversed which becomes something like “when something happened one hour later”. So if the system translates something having the similar property as above, it might not be able to recognize the pattern and reorder the words but only output the translated word by serial word order or other algorithm, and then reordering errors comes out. The fourth primary issue is tenses errors. This is very common when we analyze the translation since there is no past tense, present tense, or other tenses in Chinese. The only thing that might indicate the tense is the time specified in the

context, such as "two days ago", "one week before...", etc. Therefore, in other cases where no specific times were explicitly stated, they are probably very hard for translator to figure the correct tense out. Run-on sentence is another very common error while analyzing the translation. This problem also has two types of forms. One is that there are two complete sentences separated by comma but with no conjunction to connect these two sentences. So a very simple but useful way is to use "and" as conjunction to make these two sentences as one. And in the other situations where the two sentences have more complex relationships, we may use other conjunction to connect them instead of "and", but while analyzing this article we have not seen any other relationships beside "and". The second one is two complete sentences that may not be separated by comma can be combined into one sentence by making one sentence as clause or rephrasing them into one. The sentences in this category must have some relationship so that they can be combined. However, it seems translation system doesn't not have an efficient algorithm to detect or solve these run-on problems. Above are the five primary quality issues in the machine translation system that we found in analyzing this article translation.