LT3354 Computer-aided Translation
Assignment2

Translation quality and effort: Comparison between CAT and MT

Wan Sheung Lam (slwan9**)**
BATI
April 2017

In this assignment, I have collected 40 dish names from Café&Meal MUJI- 20 bilingual pairs as the training set to augment the translation memory (TM) and termbase (TB) for DVX2 and 20 Chinese dish names as the test set for translation into English and evaluation of the translation quality and effort required. Both the training set and test set include items from different categories, for example, 4 salad names for training set and another 4 salad names for test set.

The evaluation result reveals that CAT significantly performs better than MT in terms of translation quality. CAT scores 4.11 for average adequacy score while MT only gets 3.3, which implies that the translation from CAT contains more information than those from MT. As for average fluency score, which only considers the linguistic quality of the translation output without looking at the source text, CAT having 4.65 also wins MT having 3.55.

The superior achievement of CAT can be credited to the contribution of TM and TB. As the TM and TB are compiled from the same source with similar content, similar or same ingredient keywords enable the CAT system to match with the test set easily. The possibility of mistranslating or omitting the vocabularies can be reduced with the provision of suggestions. For instance, '魚柳' is mistranslated as incomprehensible 'cereal Liu' in MT output but 'fillet' can be used from the matched item '鱸魚柳' 'Sea Bass Fillet'. '石臼麵' which should be 'pasta' instead, even turns out to be 'sashimi' in the MT output. '炒' and '燒汁' are omitted in MT output, yet 'sauteed' and 'gravy' are recommended respectively in CAT. With the aid of matching function in CAT, the adequacy score for CAT output mostly outweighs that for MT output.

However, there are two exceptions that MT outputs having higher adequacy score- '比目魚馬鈴薯蛋沙律' and '菠菜牛蒡漢堡配季節沙律'. Translator's own knowledge towards terminologies in food domain accounts for the exceptions. The

dish name indicates the species or specific part of the ingredients. Without external aid, I cannot come up with exact translation but only general ones- 'Fish' and 'Beef' for '比目魚' and '牛蒡' respectively. Google Translate is supported by dictionaries as bases and thus able to have a more accurate translation of 'Halibut' and 'Burdock'. Actually vocabulary difficulty also bothers me in other translations. For example, '羅勒茄醬' should be 'Basil & Tomato Sauce' instead of simply 'Tomato sauce', 'Porcini' but not ambiguous 'Beef Liver Mushroom' for '牛肝菌' and 'Potato Croquette' rather than 'Potato Cake' for '薯餅'.

As for the fluency, CAT outputs receive higher scores as they consist of more organized and appropriate dish name structure that is more likely to appear in daily lives. Conjunction words, such as &, with, and in, are used to link the ingredients and indicate the main ingredient and seasoning. TM acts as reference and enables translators to follow the pattern, hence resulting in natural and fluent English translations. Compared with CAT, MT only offers the word 'with' for the translation of '配', which is less flexible. Besides, for the dish name '四品料理套餐', it gives the clue of how many dishes are included in the set and the word '品' can also be treated as a pun of '品嚐', which is translated as 'Deli' including the meaning of delicious, apart from the quantity. With the TM of '三品料理套餐', only the number needs to be changed in the suggestion. In cases like this, maybe having pun words or innovative words for attracting customers in the dish name, MT usually performs worse as machine fails to grasp the implicit meaning.

Regarding translation effort needed, CAT serves as a useful assistant and facilitates the translation process. I used fuzzy matching with score of 40 and got around 85 words matched. Although not all of them can be used, around 70% are useful after editing. Insertion is the most needed editing type as still a lot remain unmatched and untranslated. Deletion and substitution is required when the system does wrong matching, like '牛油果' was suggested separately as 'beef oil fruit' and '奶泡' as simply 'milk'. Shifting is used to rearrange the sentence order for improving the sentence structure. For instance, '甜椒醬炸雞' matched as 'pepper sauce fried chicken' and 'fried chicken' needed to be shifted forward and the word 'with' to be inserted to improve fluency. Capitalization is also needed for standardizing the format.

The training set is helpful in terms of providing unknown vocabulary suggestions, such as '味噌' and '薏米', that are not included in the original TM and TB provided on Canvas, thus increasing the number of matched items. As mentioned, it also

performs as a reference for dish name structure. However, I realize that the choice of training sets matters. If most of the inputs are unlikely to include the same ingredient like 'All Day Breakfast', 'Healthy Breakfast', and 'Pancake Set', it will just be a waste of time and effort. Ideally several items are picked from each category to increase the possibility of matching.

All in all, CAT gives more favourable translation quality than MT if appropriate and carefully selected training sets are inputted as TM and TB. Translator's knowledge also plays a key role in affecting the quality when no external aid is allowed. Despite much editing needed and time required to set up the CAT system, it is worthwhile to do good on translation effort and have translations of more publishable and comprehensible quality.

Appendix:

Café&Meal MUJI. (n.d.). Retrieved April 18, 2017, from
http://www.muji.com.hk/en/cafemeal/index.html & leaflet from branch at Festival
Walk

Training Set

青檸醬汁蝦肉芒果沙律	Shrimp & Mango Salad with Lime Dressing
玫瑰醋鮮果蘭香子沙律	Fruit & Basil Seed Salad in Rose Vinegar Dressing
意大利半風乾車厘茄沙律	Italian Semi-roasted Cherry Tomatoes
美國田園沙律	American Cobb Salad
香橙甜椒汁燴蔬菜	Stewed Vegetables with Orange & Pepper Sauce
英式蒔蘿炸魚柳配薯角	Britain Dill Batter Fish & Chips
檸檬蜜糖醬汁烤雙薯	Baked New Potato & Sweet Potato with Lemon Honey Sauce
熱情果芒果醬汁豚肉	Pork with Passion Fruit & Mango Sauce
香煎星斑柳配大根鮮果	Pan-fried Groupa Fillet with Radish & Fruit
溫泉蛋芝士烤大磨菇	Baked Portabello Mushroom with Spa Egg & Cheese
慢煮西冷牛扒配黑松露燒汁	Slow Cooked Striploin with Black Truffle Gravy
味噌烤鱸魚柳	Miso Broiled Sea Bass Fillet
牛油雞肉雜菜咖喱飯	Butter Chicken Curry Rice with Vegetables
野菌菠菜仔石臼麵	Wild Mushrooms & Baby Spinach Pasta
日本蛋沙律三文治	Japanese Egg Salad Sandwich
柚子醬汁吉列魚柳三文治	Fish Cutlet Sandwich with Yuzu Sauce
全日早餐	All Day Breakfast
三品料理套餐	3 Deli Set
兒童漢堡套餐	Kids Deli Set
雜菜薏米湯	Vegetables & Job's Tear Soup

Test set (official translation from Café&Meal MUJI)

烤春雞配牛肝菌忌廉醬汁	Roasted Spring Chicken with Porcini Cream Sauce
四品料理套餐	4 Deli Set
紫薯青蘋果沙律	Purple Sweet Potato & Green Apple Salad
櫻花蝦雜菜厚燒玉子	Sakura Shrimp & Vegetables Omelette
比目魚馬鈴薯蛋沙律	Potato & Egg Salad with Fish
牛油果薏米沙律	Avocado & Job's Tear Salad
沖繩青柑檸醬汁豚肉香草沙律	Pork & Herbs Salad with Okinawa Calamansi Dressing
意式羅勒茄醬烤茄子	Italian Eggplant Gratin with Basil & Tomato Sauce
豆腐雞肉漢堡	Tofu & Chicken Patty
精選薯餅	Selected Potato Croquette
蒜香雜菜炒牛柳粒	Sauteed Beef with Garlic & Vegetables
甜椒醬炸雞	Fried Chicken with Sweet & Chili Sauce
味噌烤比目魚柳	Baked Halibut Fillet with Miso
慢煮黑毛豬配白菌醬汁	Slow Cooked Iberico Pork with Mushroom Sauce
燒汁燴牛面頰雜菜飯	Stewed Beef Cheek & Vegetables Rice with Gravy
蔬菜肉醬石臼麵	Vegetables & Bolognese Pasta
菠菜牛蒡漢堡配季節沙律	Spinach & Burdock Burger with Seasonal Salad
芝士烤雞肉三文治	Grilled Chicken & Cheese Sandwich
兒童熱狗套餐	Kids Hot Dog Set
南瓜湯配生薑奶泡	Pumpkin Soup with Ginger Foam