AlphaMWE: Construction of Multilingual Parallel Corpora with MWE Annotations

Lifeng Han[Π], **Gareth J. F. Jones**[Π], and **Alan Smeaton**[Ω]
[Π] ADAPT Research Centre
[Ω] Insight Centre for Data Analytics
School of Computing, Dublin City University, Dublin, Ireland
lifeng.han@adaptcentre.ie {gareth.jones, alan.smeaton}@dcu.ie

Abstract

In this work, we present the construction of multilingual parallel corpora with annotation of multiword expressions (MWEs). MWEs include verbal MWEs (vMWEs) defined in the PARSEME shared task that have a verb as the head of the studied terms. The annotated vMWEs are also bilingually and multilingually aligned manually. The languages covered include English, Chinese, Polish, and German. Our original English corpus is taken from the PARSEME shared task in 2018. We performed machine translation of this source corpus followed by human post editing and annotation of target MWEs. Strict quality control was applied for error limitation, i.e., each MT output sentence received first manual post editing and annotation plus second manual quality rechecking. One of our findings during corpora preparation is that accurate translation of MWEs presents challenges to MT systems. To facilitate further MT research, we present a categorisation of the error types encountered by MT systems in performing MWE related translation. To acquire a broader view of MT issues, we selected four popular state-of-the-art MT models for comparisons namely: Microsoft Bing Translator, GoogleMT, Baidu Fanyi and DeepL MT. Because of the noise removal, translation post editing and MWE annotation by human professionals, we believe our AlphaMWE dataset will be an asset for cross-lingual and multilingual research, such as MT and information extraction. Our multilingual corpora are available as open access at github.com/poethan/AlphaMWE.

1 Introduction

Multiword Expressions (MWEs) have long been of interest to both natural language processing (NLP) researchers and linguists (Sag et al., 2002; Constant et al., 2017; Pulcini, 2020). The automatic processing of MWEs has posed significant challenges for some fields in computational linguistics (CL), such as word sense disambiguation (WSD), parsing and (automated) translation (Lambert and Banchs, 2005; Bouamor et al., 2012; Skadina, 2016; Li et al., 2019; Han et al., 2020). This is caused by both the variety and the richness of MWEs as they are used in language.

Various definitions of MWEs have included both syntactic structure and semantic viewpoints from different researchers covering syntactic anomalies, non-compositionality, non-substitutability and ambiguity (Constant et al., 2017). For instance, Baldwin and Kim (2010) define MWEs as "lexical items that: (i) can be decomposed into multiple lexemes; and (ii) display lexical, syntactic, semantic, pragmatic and/or statistical idiomaticity". However, as noted by NLP researchers for example in (Constant et al., 2017), there are very few bilingual or even multilingual parallel corpora with MWE annotations available for cross-lingual NLP research and for downstream applications such as machine translation (MT) (Johnson et al., 2016).

With regard to MWE research, verbal MWEs are a mature category that has received attention from many researchers (Maldonado et al., 2017). Verbal MWEs have a verb as the head

of the studied term and function as verbal phrases, such as "*kick* the bucket", "*cutting* capers" and "*go* to one's head". In this work, we present the construction of a multilingual corpus with vMWEs annotation, including English-Chinese, English-German and English-Polish language pairs. The same source monolingual corpus is in English with its vMWE tags from the shared task affiliated with the SIGLEX-MWE workshop in 2018 (Walsh et al., 2018; Ramisch et al., 2018). Several state-of-the-art (SOTA) MT models were used to perform an automated translation, and then human post editing and annotation for the target languages was conducted with cross validation to ensure the quality, i.e., with each sentence receiving post-editing and rechecking by at least two people.

In order to get a deeper insight into the difficulties of processing MWEs we carried out a categorisation of the errors made by MT models when processing MWEs. From this we conclude that current state-of-the-art MT models are far from reaching parity with humans in terms of translation performance, especially on idiomatic MWEs, even for sentence level translation, although researchers sometimes claim otherwise (Wu et al., 2016; Hassan et al., 2018).

The rest of this paper is organised as follows. In the next section we present related work and then detail the corpus preparation stages including selection of MT models. We then look at the various kinds of issues that MT has with MWEs. This analysis, along with the public release of the corpora as a resource to the community, is the main contribution of the paper.

2 Related Work

There are a number of existing studies which focus on the creation of *monolingual* corpora with vMWE annotations, such as the PARSEME shared task corpora (Savary et al., 2017; Ramisch et al., 2018). The 2020 edition of this task covers 14 languages including Chinese, Hindi, and Turkish as non-European languages. Some work from monolingual English corpora includes the MWE aware "English Dependency Corpus" from the Linguistic Data Consortium (LDC2017T01) that covers *compound words* used to train parsing models. Also related to this are English MWEs from "web reviews data" by Schneider et al. (2014) that covers *noun, verb* and *preposition super-senses* and English verbal MWEs from Walsh et al. (2018) and Kato et al. (2018) that covers PARSEME shared task defined vMWE categories. However, all these works were performed in monolingual settings, independently by different language speakers without any bilingual alignment. These corpora are helpful for monolingual MWE research such as *discovery* or *identification*, however, it would be difficult to use these corpora for bilingual or multilingual research such as MT or cross-lingual information extraction.

The work most related to ours is from Vincze (2012), who created an English-Hungarian parallel corpus with annotations for light verb constructions (LVCs). As many as 703 LVCs for Hungarian and 727 for English were annotated in this work, and a comparison between English and Hungarian data was carried out. However, the work did not cover other types of vMWEs, for instance inherently adpositional verbs, verbal idioms, or verb-particle constructions, and it was not extended to any other language pairs. In our work, we annotate in a multilingual setting including far distance languages such as English, German, Polish and Chinese, in addition to the extension of vMWE categories. In other recent work Han et al. (2020), we performed an automatic construction of bilingual MWE terms based on a parallel corpus, in this case English-Chinese and English-German. We first conducted automated extraction of monolingual MWEs based on part-of-speech (POS) patterns and then aligned the two side monolingual MWEs into bilingual terms based on statistical lexical translation probability. However, due to the automated procedure, the extracted bilingual "MWE terms" contain not only MWEs but also normal phrases. Part of the reason for this is due to the POS pattern design which is a challenging task for each language and needs to be further refined (Skadina, 2016; Rikters and Bojar, 2017; Han et al., 2020).

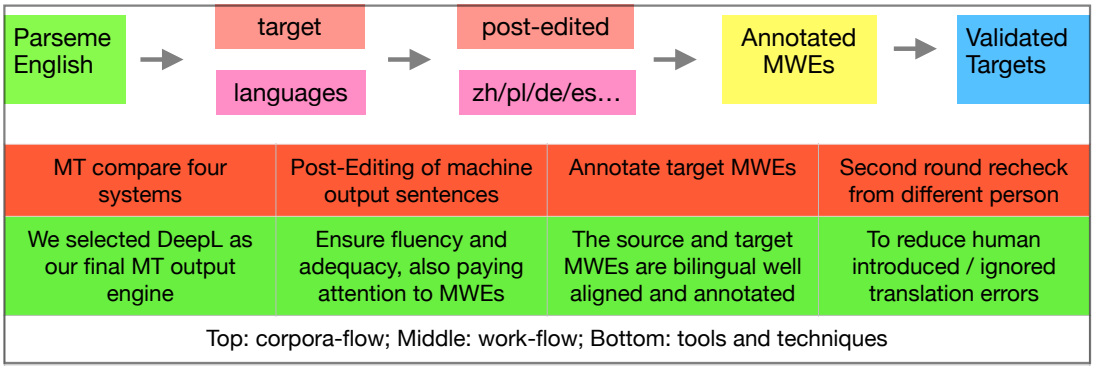

Figure 1: Workflows to prepare AlphaMWE.

3 Experimental Work

In this section we describe our corpus preparation method, selection of the MT models used in our investigation, and the resulting open-source corpora AlphaMWE.

3.1 Corpus Preparation

To construct a well aligned multilingual parallel corpus, our approach is to take a monolingual corpus from the PARSEME *vMWE discovery and identification shared task* as our root corpus. Our rationale here is that this shared task is well established and its process of tagging and categorisation is clear. Furthermore, as we plan to extend the MWE categories in the future, we enrich the PARSEME shared task corpus with potential for other downstream research and applications, including bilingual and multilingual NLP models. The English corpus (Walsh et al., 2018) we used from the PARSEME shared task follows the annotation guidelines having a broad range of vMWE categories tagged. These include inherently adpositional verbs, light verb constructions, multi-verb constructions, verbal idioms, and verb-particle constructions. The English corpus contains sentences from several different topics, such as news, literature, and IT documents. For the IT document domain, vMWEs are usually easier or more straightforward to translate, with a high chance of repetition, e.g. “apply filter” and “based on”. For the literature annotations, the vMWEs include richer samples with many idiomatic or metaphor expressions, such as “cutting capers” and “gone slightly to someone’s head” that cause MT issues.

Fig. 1 shows our workflow. This first used MT models to perform automated translation for the target language direction, then human post editing of the output hypotheses with annotation of the corresponding target side vMWEs which are aligned with the source English ones. Finally, to avoid human introduced errors, we apply a cross validation strategy, where each sentence receives at least a second person’s quality checking after the first post-editing. Tagging errors are more likely to occur if only one human has seen each sentence (we discuss some error samples from English source corpus in later sections).

3.2 MT Model Selection

We tested a number of example sentences from the English testset to compare state-of-the-art MT from Microsoft Bing (Chowdhary and Greenwood, 2017), GoogleMT (Vaswani et al., 2017), Baidu Fanyi (Sun et al., 2019), and DeepL[1], as in Fig. 2. We illustrate the comparative performances with two worked example translations. As a first example sentence, GoogleMT and Bing Translator have very similar outputs, where the MT output sentences try to capture and produce as much information as possible, but make the sentences redundant or awkward to read, such as the phrase “验证... 是否验证了 (yàn zhèng ... Shì fǒu yàn zhèng le)” where they use a repeated word “验证” (yàn zhèng, *verify*). Although the DeepL Translator does not produce a

[1]https://www.deepl.com/en/translator (All testing was performed in 2020/07 from 4 MT models)

Two sample sentences' MT outputs comparison from head of test file	
Source	# text = SQL Server verifies that the account name and password were validated when the user logged on to the system and grants access to the database, without requiring a separate logon name or password.
DeepL	# text = SQL Server 会在用户登录系统时验证账户名和密码，并授予对数据库的访问权限，而不要求单独的登录名或密码。
Google	# text = SQL Server验证用户登录系统时是否验证了帐户名和密码，并授予对数据库的访问权限，而无需单独的登录名或密码。
Bing	[text] SQL Server 验证用户登录到系统时是否验证了帐户名称和密码，并授予对数据库的访问权限，而无需单独的登录名称或密码。
Baidu	#或者，在没有密码的情况下验证用户名和用户名是否被登录到数据库中，并且系统是否授予登录权限。
Ref.	# 文本 = SQL Server 会在用户登录系统时验证账户名和密码的有效性，并授予对数据库的访问权限，而不要求单独的登录名或密码。
Source	# text = See the http://officeupdate.microsoft.com/, Microsoft Developer Network Web site for more information on TSQL.
DeepL	# text = 有关 TSQL 的更多信息，请参见 http://officeupdate.microsoft.com/，Microsoft Developer Network Web 站点。
Google	# text =有关Microsoft SQL的更多信息，请参见http://officeupdate.microsoft.com/，Microsoft开发人员网络网站。
Bing	[文本] 有关 TSQL http://officeupdate.microsoft.com/了解有关 TSQL 的相关信息，请参阅 Microsoft 开发人员网络网站。
Baidu	#text=参见http://officeupdate.microsoft.com/，以获取有关TSQL的详细信息。
Ref.	# text = 有关 TSQL 的更多信息，请参见 http://officeupdate.microsoft.com/，微软开发人员网络互联网站点。
Blue: redundancy; green: adding error; pink: reordering error; yellow: dropping error.	

Figure 2: Sample comparison of outputs from four MT models.

perfect translation since it drops the source word "validated" which should be translated as "有效性 (yǒu xiào xìng)"(as one candidate translation), the overall output is fluent and the source sentence meaning is mostly preserved. Baidu translator yields the worst output in this example. It produces some words that were not in the source sentence (或者, huò zhě, *or*), loses some important terms' translation from source sentence ("SQL Server", the subject of the sentence), and the reordering of the sentence fails resulting in an incorrect meaning ("在没有密码的情况下, zài méi yǒu mì mǎ de qíng kuàng xià" is moved from the end of the sentence to the front and made as a condition). So, for this case, DeepL performed best.

For a second example sentence, GoogleMT confused the original term TSQL as SQL. Bing MT had a similar issue with the last example, i.e. it produced redundant information "有关 (yǒu guān)" (*about/on*). In addition it concatenated the website address and normal phrase "了解有关 (liǎo jiě yǒu guān)" together with a hyperlink. GoogleMT and Bing both translate half of the source term/MWE "Microsoft Developer Network Web" as "Microsoft 开发人员网络网站" (kāi fā rén yuán wǎng luò wǎng zhàn) where they kept "Microsoft" but translated "Developer Network Web". Although this is one reasonable output since Microsoft is a general popular named entity while "Developer Network Web" consists of common words, we interpret "Microsoft Developer Network Web" as a named entity/MWE in the source sentence that consists of all capitalised words which would be better translated overall as "微软开发人员网络网站 (wēi ruǎn kāi fā rén yuán wǎng luò wǎng zhàn)" or be kept as the original capitalised words as a foreign term in the output, which is how DeepL outputs this expression. However, Baidu totally drops out this MWE translation and another word translation is not accurate, "more" into 详细 (xiáng xì). Based on these samples, we chose to use DeepL as the provider of our MT hypotheses.

3.3 Result: AlphaMWE

Regarding the size of the corpus, we extracted all 750 English sentences which have vMWE tags included. The target languages covered so far include Chinese, German and Polish with sample sentences in Appendix (Fig. 11). There are several situations and decisions that are worth noting: a) when the original English vMWEs are translated into a general phrase in the target language but not choosing sequence of MWEs, we tried to offer two different references, with

Source	At the corner of 72nd Street and Madison Avenue, he waved down a cab.
DeepL	在72街和麦迪逊大道的拐角处，他向一辆出租车招手。 Zài 72 jiē hé mài dí xùn dà dào de guǎi jiǎo chù, tā xiàng yī liàng chū zū chē zhāo shǒu.
Bing	在72街和麦迪逊大道的拐角处，他挥手示意一辆出租车。 zài 72 jiē hé mài dí xùn dà dào de guǎi jiǎo chù , tā huī shǒu shì yì yí liàng chū zū chē.
Google	在第72街和麦迪逊大街的拐角处，他挥舞着出租车。 Zài dì 72 jiē hé mài dí xùn dà jiē de guǎi jiǎo chù, tā huī wǔ zhe chū zū chē.
Baidu	在72街和麦迪逊大街的拐角处，他挥手叫了一辆出租车。 zài 72 jiē hé mài dí xùn dà jiē de guǎi jiǎo chù, tā huī shǒu jiào le yí liàng chū zū chē.
Ref.	在72街和麦迪逊大道的拐角处，他招手示停了一辆出租车。 Zài 72 jiē hé mài dí xùn dà dào de guǎi jiǎo chù, tā zhāo shǒu shì tíng le yī liàng chū zū chē.

Figure 3: MT issues with MWEs: common sense. Pinyin offered by GoogleMT with post-editing.

one of them being revised in a vMWE/MWE presentation in the target; b) when the original English sentence terms were translated into the correct target language but in a different register, e.g. the source language has low register (thx, for instance), we offer two reference sentences, with one of them using the same low register and the other with (formal) full word spelling; c) for the situations where a single English word or normal phrase is translated into a typical vMWE in the target language, or both source and target sentences include vMWEs but the source vMWE was not annotated in the original English corpus, we made some additions to include such vMWE (pairs) into AlphaMWE; d) for some wrong/incorrect annotation in the source English corpus, or some mis-spelling of words, we corrected them in AlphaMWE; e) we chose English as root/source corpus, since the post-editing and annotation of target languages requires the human annotators to be fluent/native in both-side languages, and all editors were fluent in English as well as being native speakers in the specific target languages respectively.

We examined the development and test data sets from the annual Workshop of MT (WMT) (Bojar et al., 2017) and also from the NIST MT challenges where they offered approximately 2,000 sentences for development/testing over some years. This means that our bilingual/multilingual corpora with 750 sentences is comparable to such standard shared task usage.

4 MT Issues with MWEs

We performed an analysis of the behaviour of various MT systems when required to translate MWEs or MWEs related context. Due to space limitations, in this paper we focus on the English→Chinese language pair. We also highlight some issues on English→German and English→Polish in the next section, but leave the detailed analysis of other language pairs for future work. When MT produces incorrect or awkward translations this can fall into many different categories, and from our analysis we classify them as: *common sense, super sense, abstract phrase, idiom, metaphor and ambiguity*, with ambiguity further sub-divided. These classifications are to be further refined in the future, e.g. the differences between metaphor and idiom are usually fuzzy. We now list each of these with examples to support future MT research on improving the quality of MT when handling MWEs.

4.1 Common Sense

The first error category is the *common sense* issue. For instance, the sentence in Fig. 3 includes the vMWE "waved down" which in general understanding indicates that "he succeeded in getting the cab" and not only "waved his hand". However, in the translation by DeepL and Bing this vMWE was wrongly translated as "he waved his hand to the cab" missing part of the original meaning; the MT output by GoogleMT is also incorrect, saying "he waves with the cab in hand";

Source	Each time he took a walk, he felt as though he were leaving himself behind, and by giving himself up to the movement of the streets, by reducing himself to a seeing eye, he was able to escape the obligation to think, and this, more than anything else, brought him a measure of peace, a salutatory emptiness within.
DeepL	每一次散步，他都觉得自己好像把自己抛在了脑后，把自己交给了街道的运动，把自己还原成了一只看得见的眼睛，他就能摆脱思考的义务，这比什么都重要，这给他带来了一定程度的平静，内心的空虚让人敬佩。(jiāo gěi; nèi xīn de kōng xū ràng rén jìng pèi)
Bing	每次散步时，他都觉得自己好像要把自己抛在脑后，通过向街道运动自在，通过把自己留眼，他能够逃避思考的义务，这比什么都重要，给他带来了一定程度的和平，一种有益的空虚。(zì zài; yǒu yì de kōng xū)
Google	每次散步时，他都会感觉自己好像在抛弃自己，投身于大街小巷，睁大了眼睛，这使他摆脱了思考的义务，而这， 最重要的是，他带给他一种和平的感觉，一种内在的称呼空虚。(tóu shēn yú; nèi zài de chēng hu kōng xū)
Baidu	每次他散步，他都觉得自己好像是在离开自己，把自己交给街道的流动，把自己变成一个有眼光的人，他就可以逃避思考的义务，而这比其他任何事情都能给他带来某种程度的平和，一种内在的致意的空虚。(nèi zài de zhì yì de kōng xū)
Ref.	每一次散步，他都觉得自己好像把自己抛在了脑后，投身于大街小巷，把自己还原成了一只看得见的眼睛，他就能摆脱思考的义务，这给他带来了某种程度的平静和内心悦纳的空无，远胜于其他。(nèi xīn yuè nà de kōng wú)

Figure 4: MT issues with MWEs: super sense.

Source	Quinn had his doubts, but this was all he had to **go on**, his only bridge to the present.
DeepL	奎恩有他的疑虑，但这是他唯一的依靠，是他通往现在的唯一桥梁。(wéi yī de yī kào)
Bing	奎因有他的怀疑，但这就是他必须**去**的，他唯一的桥梁到现在。 (bì xū qù de)
Google	奎因有他的疑惑，但这就是他所要**做**的，是他通往现在的唯一桥梁。 (suǒ yào zuò de)
Baidu	奎因有他的疑虑，但这是他所要**做**的，是他通往现在的唯一桥梁。 (suǒ yào zuò de)
Ref.	奎恩曾有他的疑虑，但这是他**开展工作**的所有依据，是他通往现在的唯一桥梁。(**kāi zhǎn gōng zuò** de suǒ yǒu yī jù)

Figure 5: MT issues with MWEs: abstract phrase

the Baidu translation of this sentence is semantically correct that "he waved and got one cab" though it does not use a corresponding Chinese side vMWE " 招手示停 (zhāo shǒu shì tíng)"[2].

4.2 Super Sense

For this category of translation issue, it is related to *a form of state of mind* and we need to make a logical prediction to guess the *positiveness* or *negativeness* of some words, in the choice of Chinese characters. As in Fig. 4, the MT systems each have advantages for different parts of this long sentence. However, none of them is perfect. For instance, for the translation of vMWE "giving (himself) up (to)", the DeepL and Baidu outputs give very literal translation by saying "he gives himself to", the Bing translator drops the vMWE, while GoogleMT preserves the correct meaning in the translation "投身于 (tóu shēn yú)" from the reference indicating "he devoted himself". However, GoogleMT's output for the phrase "salutatory emptiness within" is very poor and makes no sense; the reference is "the emptiness that he welcomes" for which Baidu has a closer translation "内在的致意的空虚 (nèi zài de zhì yì de kōng xū)". All four MT outputs also use the same Chinese words "空虚 (kōng xū)" which is a term with negative meaning, however, the sentence indicates that he is welcoming this emptiness, which should be the corresponding Chinese words "空无 (kōng wú)", an unbiased or positive meaning.

4.3 Abstract Phrase

The *abstract phrases* can have different exact meanings and we usually need some background information from the sentence or paragraph to select the correct word choices in the target

[2]We give full sentence pronunciation (in Pinyin) of Chinese characters in this figure, for the following examples, we only present the Chinese Pinyin for MWEs and studied words of the sentences to save space.

Source	I was smoking my pipe quietly by my dismantled steamer, and saw them all cutting capers in the light, with their arms lifted high, when the stout man with mustaches came tearing down to the river, a tin pail in his hand, assured me that everybody was 'behaving splendidly, splendidly, dipped about a quart of water and tore back again.
DeepL	我在拆开的汽船旁静静地抽着烟斗，看到他们都在灯光下高举双臂切着荠菜，这时，那个留着胡子的壮汉撕扯着来到河边，手里拿着一个铁皮桶，向我保证说，大家都'表现得很精彩，很精彩，浸了大约一夸脱的水，又撕扯回来了。 (qiè zhe jì cài; sī chě huí lái)
Bing	我悄悄地用我拆除的蒸笼抽我的烟斗，看到他们所有的切割披肩在光中，他们的手臂抬起来高，当粗壮的人用胡子来到河中，一个锡桶在他的手中，向我保证，每个人都'表现辉煌，辉煌，浸约夸脱的水，并再次撕毁。(qiē gē pī jiān; zài cì sī huǐ)
Google	我正用拆卸的蒸锅安静地抽着烟斗，看到他们都在灯光下割下刺山柑，双臂高高举起，当时那留着小胡子的矮壮男人向河边撕下，手里拿着一个铁桶，向我保证 每个人都表现得非常出色，表现得非常出色，蘸了一夸脱的水，然后又撕了回来。(gē xià cì shān gān; sī le huí lái)
Baidu	我正静静地在我拆掉的汽船旁抽着烟斗，看见他们都高举着胳膊，在灯光下割着山柑，这时那个长着胡子的胖男人手里拿着一个锡桶，朝河里跑来，向我保证每个人都表现得很好，很漂亮，蘸了一夸脱水，然后又往回跑。(gē zhe shān gān; wǎng huí pǎo)
Ref.	我在拆开的汽船旁静静地抽着烟斗，看到他们都在灯光下欢呼雀跃，高举双臂，这时，那个留着胡子的大块头，手里拿着一个铁皮桶，快速来到河边，向我确保大家都"表现得很精彩，很精彩"，他浸了大约一夸脱的水，又快速回去了。(huān hū què yuè; kuài sù huí qù)

Figure 6: MT issues with MWEs: idioms

language[3]. With the example sentence in Fig. 5, from the context, we know that "go on" in this sentence means "to work from" using all the information he had. The phrase "this was all he had to go on" is then to be interpreted as "this is all the information he had to work from". At the end of the sentence, "the present" is the "present person" he needs to look for (with the picture of this person's younger age portrait). However, Bing translated it as "this is (where) he had to go" which is an incorrect interpretation of "had to go"; furthermore, Bing's translation of the second half of the sentence kept the English order, without any reordering between the words, which is grammatically incorrect in Chinese, i.e. "他唯一的桥梁到现在 (tā wéi yī de qiáo liáng dào xiàn zài)". GoogleMT and Baidu translated it as "what he need to do" which is also far from correct, while DeepL successfully translated the part "his only thing to relying on" but dropped the phrase "go on", i.e., *to do what*. *Abstract Phrase* can include *Super Sense* as its sub-category, however, it does not necessarily relate to a state of mind.

4.4 Idioms

The use of *idioms* often causes wrongly translated sentences, mostly resulting in humorous output due to literal translation. For example, in the sentence in Fig. 6, the vMWEs "*cutting capers*" and "tore back" are never translated correctly at the same time by any of the four MT models we used. The idiom "cutting capers" indicates frolic or romp, to "act in the manner of a young goat clumsily frolicking about" and here it means "they are in a happy mood, playful and lively movement" which should properly be translated as the corresponding Chinese idiom "欢呼雀跃 (huān hū què yuè, happily jumping around like sparrows)". However, all four MT models translated it literally into "cutting" actions just with different subjects, i.e., what they cut. The idiom (slang) "tore back" means the stout man *walked back rapidly*, which the Baidu translation gives the closest translation as "往回跑 (wǎng huí pǎo, run back)" but the other three models translated into an action "tear something (to be broken)" which is incorrect.

4.5 Metaphor

The first sentence vMWE "blown to bits" in Fig. 7 is a *metaphor* to indicate "everything is gone", instead of the physical "blowing action". However, the three MT models DeepL, GoogleMT and

[3]it sometimes belongs to the context-unaware ambiguity (CUA) that we will mention later, however, CUA not necessarily means "abstract phrase", and usually needs paragraph information, not only sentence level. Furthermore, in some situations, we just don't know how to interpret "abstract phrase", i.e. the candidate interpretations are unknown without context, and this is different from ambiguity.

Source	The what? Auster laughed, and in that laugh everything was suddenly blown to bits. The chair was comfortable, and the beer had gone slightly to his head.
DeepL	那个什么？奥斯特笑了，在这笑声中，一切突然被炸得粉碎。(bèi zhà dé fěn suì) 椅子很舒服，啤酒已经微微到了他的头上。 (wēi wēi dào le tā de tóu shàng)
Bing	什么？奥斯特笑了，在笑，一切都突然被吹成位。(bèi chuī chéng wèi) 椅子很舒服，啤酒稍微到他的头去了。 (shāo wēi dào tā de tóu qù le)
Google	什么啊 Auster笑了起来，在那笑声中，一切突然被炸碎了。 (bèi zhà suì le) 椅子很舒服，啤酒微微飘到他的头上。(wēi wēi piāo dào tā de tóu shàng)
Baidu	什么？奥斯特笑了，在那笑声中，一切都突然被炸成碎片。(bèi zhà chéng suì piàn) 椅子很舒服，啤酒已经稍稍流到他的头上了。(shāo shāo liú dào tā de tóu shàng le)
Ref.	那个什么？奥斯特笑了，在这笑声中，一切突然化为乌有。(huà wéi wū yǒu) 椅子很舒服，啤酒已经微微让他上了头。 (wēi wēi ràng tā shàng le tóu)

Figure 7: MT issues with MWEs: metaphor

Source	But it did not give me the time of day.
DeepL	但它并没有给我时间。 (gěi wǒ shí jiān)
Bing	但它没有给我一天的时间。 (gěi wǒ yī tiān de shí jiān)
Google	但这没有给我一天的时间。 (gěi wǒ yī tiān de shí jiān)
Baidu	但它没有给我一天中的时间。 (gěi wǒ yī tiān zhōng de shí jiān)
Ref.	但我没有感到这个对于我特殊 / 但这不是我的菜。(gǎn dào zhè ge duì yú wǒ tè shū / … wǒ de cài)
Context	An old Mormon missionary in Nauvoo once gripped my knee hard as we sat side by side, and he put his arm about me and called me "Brother." We'd only met ten minutes before. He took me to his good bosom. His eyes began to mist. I was a prospect, an exotic prospect in old tennis shoes and a sweatshirt. His heart opened to me. It opened like a cuckoo clock. But it did not …

Figure 8: MT issues with MWEs: context-unaware ambiguity

Baidu translate it as "exploded into pieces (by bombs)", while BingMT translates it even more literally into "blown to (computer) bits". There is a corresponding Chinese vMWE "化为乌有 (huà wéi wū yǒu, vanish into nothing)" which would be a proper choice for this source vMWE translation. The second sentence vMWE "gone (slightly) to his head" is a metaphor to indicate "got slightly drunk". However, all four MT models translate it as physically "beer moved to his head" but by slightly different means such as *flow* or *flutter*. The corresponding translation as a MWE should be "微微让他上了头 (wéi wéi ràng tā **shàng le tóu**)", using the same characters, but the character order here makes so much difference, meaning "slightly drunk".

4.6 Ambiguity

We encountered different kinds of situation that cause ambiguity in the resulting translation when it meets MWEs or named entities, so we further divide ambiguity into three sub-classes.

4.6.1 Context-Unaware Ambiguity

In this case, the *context*, i.e. the background information, is needed for correct translation of the sentence. For instance, see Fig. 8. DeepL gives the translation "it did not give me time though", while Bing and GoogleMT give the same translation "it/this did not give me one day's time" and Baidu outputs a grammatically incorrect sentence. From the pre-context, we understand that it means the speaker "did not feel that is special to him" or "did not have affection of that" after *all the Mormon missionary's effort towards him.* Interestingly, there is a popular Chinese idiom (slang) that matches this meaning very well "不是我的菜 (bù shì wǒ

Source	The moment they know the de-gnoming's going on they storm up to have a look. Then someone says that it can't be long now before the Russians write Arafat off.
DeepL	他们一知道去核的事，就会冲上去看一看。 (qù hé) 然后有人说，现在用不了多久，俄罗斯人就会**把**阿拉法特**注销**。(**bǎ** ā lā fǎ tè **zhù xiāo**)
Bing	当他们知道去诺格明是怎么回事， 他们冲了起来看看。 (qù nuò gé míng) 然后有人说，现在俄罗斯人要不长了，就**把**阿拉法特**注销**了。(**bǎ** ā lā fǎ tè **zhù xiāo** le)
Google	当他们知道正在逐渐消失的那一刻，他们便冲上去看看。 (zhèng zài zhú jiàn xiāo shī) 然后有人说，不久之后俄罗斯人**将**阿拉法特**注销**。 (**jiāng** ā lā fǎ tè **zhù xiāo**)
Baidu	他们一知道德格诺明正在进行，就冲上去看一看。 (dé gé nuò míng) 然后有人说，俄国人很快就会把阿拉法特**一笔勾销**了。 (bǎ ā lā fǎ tè **yī bǐ gōu xiāo** le)
Ref.	一知道去地精的事在进行，他们就冲上去观看。 (qù dì jīng) 然后有人说，现在用不了多久，俄罗斯人就会**把**阿拉法特**下課** / 让…下台。 (**bǎ** ā lā fǎ tè **xià kè;** ràng…xià tái)

Figure 9: MT issues with MWEs: social/literature-unaware ambiguity

de cài, literally *not my dish*)". From this point of view, the context based MT model deserves some more attentions, instead of only focusing on sentence level. When we tried to put all background context information as shown in Fig.8 into the four MT models, they produce as the same output for this studied sentence, as for sentence level MT. This indicates that current MT models still focus on sentence-by-sentence translation when meeting paragraphs, instead of using context inference.

4.6.2 Social/Literature-Unaware Ambiguity

In this case, *social knowledge* of current affairs from news, or *literature knowledge* about some newly invented entities / phrases are required in order to get a correct translation output. For instance, Fig. 9 includes two sentences, one from politics and another from literature.

In the first sentence, "de-gnoming" is a literature word from Harry Potter, invented by its author, to refer to the process of ridding a garden of gnomes, *a small magical beast*. Without this literature knowledge it is not possible to translate the sentence correctly. For instance, even though this sentence is from a very popular novel that has been translated into most languages, DeepL translated it as "去核 (qù hé, de-nuclear)", Bing translated it as "去诺格明 (qù nuò gé míng, *de*-nuògémíng" where "nuògémíng" is a simulation of the pronunciation of "gnoming" in a Chinese way, Baidu translated it as "德格诺明 (dé gé nuò míng)" which is the simulation of the pronunciation of the overall term "de-gnoming".

In the second sentence, "write Arafat off" is to dismiss "Yasser Arafat", Chairman of the Palestine Liberation Organization, who is a historical person's name. However, all three models DeepL, Bing, and GoogleMT translated it into "把/将阿拉法特注销 (bǎ/jiāng ā lā fǎ tè zhù xiāo, *deregister Arafat*)" which treated "Arafat" as a tittle of certain policy/proceeding, not being able to recognize it as a personal named entity, while Baidu made the effort to use the Chinese idiom "一笔勾销 (yī bǐ gōu xiāo, *cancel everything*, or *never mention historical conflicts*)" for "write off" but it is not a correct translation. Interestingly, if we put these two sentences into a web search engine it retrieves the correct web pages as context in the top list of the search result. This may indicate that future MT models could consider to include web search results as part of their knowledge of background for translation purposes.

4.6.3 Coherence-Unaware Ambiguity

This kind of MWE ambiguity can be solved by the *coherence* of the sentence itself, for instance, the example in Fig. 10. The four MT models all translated the vMWE itself "have an operation" correctly in meaning preservation by "做/接受/动手术 (zuò/jiē shòu/dòng shǒu shù)" just with different Chinese word choices. However, none of the MT models translated the "reason of the operation", i.e., "complaint" correctly. The word complaint has two most commonly

Source	Two months ago I had to have an operation for a serious ***complaint***.
DeepL	两个月前，我因为一次严重的**投诉**不得不做手术。(**tóu sù** … zuò shǒu shù)
Bing	两个月前，我不得不做一个严重的**投诉**手术。 (zuò … **tóu sù** shǒu shù)
Google	两个月前，我不得不接受一次手术以应对严重的**投诉**。 (jiē shòu yī cì shǒu shù … **tóu sù**)
Baidu	两个月前，我因为严重的**投诉**不得不动手术。 (**tóu sù** … dòng shǒu shù)
Ref.	两个月前，我因为一次严重的**症状**不得不做手术。(**zhèng zhuàng** … zuò shǒu shù)

Figure 10: MT issues with MWEs: coherence-unaware ambiguity

used meanings "a statement that something is unsatisfactory or unacceptable" or "an illness or medical condition" and all four models chose the first one. According to simple logic of social life, people do not need to "have an operation" due to "a statement", instead their "medical condition" should have been chosen to translate the word "complaint". Because of the incorrectly chosen candidate translation of the word "complaint", Bing's output even invented a new term in Chinese "投诉手术 (tóu sù shǒu shù, *a surgery of complaint statement kind*)" which makes no sense.

5 Conclusion and Future Work

In this paper, we presented the construction of multilingual parallel corpora, AlphaMWE, with vMWEs as pioneer annotations by native speakers of the corresponding languages. We described the procedure of MT model selection, human post editing and annotation, and compared different state-of-the-art MT models and classified the MT errors from vMWEs related sentence/context translations. We characterised the errors into different categories to help MT research to focus on one or more of them to improve the performance of MT.

We performed the same process as described here for English→Chinese, English→German and English→Polish and similarly categorised the MT issues when handling MWEs. The English→German issues can be categorized into: (1) there are cases where the corresponding German translation of English MWEs can be one word, which is partially because that German has separable verbs, (2) the automated translation to German is biased towards choosing the polite or formal form of the words, which is generally fine but depends on the context to decide which form is more suitable, and (3) English vMWEs are often not translated as vMWEs to German. In the main, English→Polish MT errors fall into the category of coherence-unaware errors, literal translation errors and context unaware situation errors.

We name our process as AlphaMWE to indicate that we will continue to maintain the developed corpora which are publicly available and extend them into other possible language pairs, e.g. Spanish, French and Italian (under-development). We also plan to extend the annotated MWE genres beyond the vMWEs defined in the PARSEME shared task.

Acknowledgements

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund. The input of Alan Smeaton is part-funded by SFI under grant number SFI/12/RC/2289 (Insight Centre). The authors are very grateful to their colleagues who helped to create the AlphaMWE corpora by post editing and annotation work across all language pairs, to Yi Lu and Dr. Paolo Bolzoni for helping with the experiments, to Lorin Sweeney, Roise McGagh, and Eoin Treacy for discussions about English MWEs and terms, Yandy Wong for discussion of Cantonese examples, and Hailin Hao and Yao Tong for joining the first discussion. The authors also thank the anonymous reviewers for their thorough reviews and insightful comments.

References

Timothy Baldwin and Su Nam Kim. 2010. Multiword expressions. In *Handbook of Natural Language Processing, Second Edition*, pages 267–292. Chapman and Hall.

Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Shujian Huang, Matthias Huck, Philipp Koehn, Qun Liu, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Raphael Rubino, Lucia Specia, and Marco Turchi. 2017. Findings of the 2017 conference on machine translation (WMT17). In *Proceedings of the Second Conference on Machine Translation*, pages 169–214, Copenhagen, Denmark, September. Association for Computational Linguistics.

Dhouha Bouamor, Nasredine Semmar, and Pierre Zweigenbaum. 2012. Identifying bilingual multi-word expressions for statistical machine translation. In *Conference on Language Resources and Evaluation*.

Vishal Chowdhary and Scott Greenwood. 2017. Emt: End to end model training for msr machine translation. In *Proceedings of the 1st Workshop on Data Management for End-to-End Machine Learning*.

Mathieu Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner, and Amalia Todirascu. 2017. Survey: Multiword expression processing: A Survey. *Computational Linguistics*, 43(4):837–892.

Lifeng Han, Gareth J. F. Jones, and Alan Smeaton. 2020. MultiMWE: Building a multi-lingual multiword expression (MWE) parallel corpora. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 2970–2979, Marseille, France. European Language Resources Association.

Hany Hassan, Anthony Aue, C. Chen, Vishal Chowdhary, J. Clark, C. Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, W. Lewis, M. Li, Shujie Liu, T. Liu, Renqian Luo, Arul Menezes, Tao Qin, F. Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and M. Zhou. 2018. Achieving human parity on automatic chinese to english news translation. *ArXiv*, abs/1803.05567.

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda B. Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's multilingual neural machine translation system: Enabling zero-shot translation. *CoRR*, abs/1611.04558.

Akihiko Kato, Hiroyuki Shindo, and Yuji Matsumoto. 2018. Construction of Large-scale English Verbal Multiword Expression Annotated Corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018.

Patrik Lambert and Rafael E. Banchs. 2005. Data Inferred Multi-word Expressions for Statistical Machine Translation. In *Proceedings of Machine Translation Summit X*, pages 396–403, Thailand.

Xiaoqing Li, Jinghui Yan, Jiajun Zhang, and Chengqing Zong. 2019. Neural name translation improves neural machine translation. In Jiajun Chen and Jiajun Zhang, editors, *Machine Translation*, pages 93–100, Singapore. Springer Singapore.

Alfredo Maldonado, Lifeng Han, Erwan Moreau, Ashjan Alsulaimani, Koel Dutta Chowdhury, Carl Vogel, and Qun Liu. 2017. Detection of verbal multi-word expressions via conditional random fields with syntactic dependency features and semantic re-ranking. In *The 13th Workshop on Multiword Expressions @ EACL 2017*. ACL.

Virginia Pulcini. 2020. English-derived multi-word and phraseological units across languages in the global anglicism database. *Textus, English Studies in Italy*, (1/2020):127–143.

Carlos Ramisch, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archna Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoa Iñurrieta, Jolanta Kovalevskaitė, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya, and Abigail Walsh. 2018. Edition 1.1 of the PARSEME shared task on automatic identification of verbal multiword expressions. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 222–240, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Matīss Rikters and Ondřej Bojar. 2017. Paying Attention to Multi-Word Expressions in Neural Machine Translation. In *Proceedings of the 16th Machine Translation Summit*, Nagoya, Japan.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for nlp. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 1–15, Berlin, Heidelberg. Springer Berlin Heidelberg.

Agata Savary, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova, and Antoine Doucet. 2017. The PARSEME shared task on automatic identification of verbal multiword expressions. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 31–47, Valencia, Spain.

Nathan Schneider, Spencer Onuffer, Nora Kazour, Emily Danchik, Michael T. Mordowanec, Henrietta Conrad, and Noah A. Smith. 2014. Comprehensive annotation of multiword expressions in a social web corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, pages 455–461, Reykjavik, Iceland, May. European Languages Resources Association.

Inguna Skadina. 2016. Multi-word expressions in english-latvian machine translation. *Baltic J. Modern Computing*, 4:811–825.

Meng Sun, Bojian Jiang, Hao Xiong, Zhongjun He, Hua Wu, and Haifeng Wang. 2019. Baidu neural machine translation systems for WMT19. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 374–381, Florence, Italy. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Conference on Neural Information Processing System*, pages 6000–6010.

Veronika Vincze. 2012. Light verb constructions in the SzegedParalellFX English–Hungarian parallel corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2381–2388, Istanbul, Turkey, May. European Language Resources Association.

Abigail Walsh, Claire Bonial, Kristina Geeraert, John P. McCrae, Nathan Schneider, and Clarissa Somers. 2018. Constructing an annotated corpus of verbal MWEs for English. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 193–200, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144.

Appendix A: AlphaMWE Corpus Presentation Examples.

As shown in the examples (Fig. 11) from Chinese, German and Polish, all involved languages are sentence by sentence aligned, including the vMWEs paired with order which are put behind the sentences into the bracket pairs. AlphaMWE also includes statistics of the annotated vMWEs, and a multilingual vMWEs glossary. The AlphaMWE corpora are divided evenly into five portions which were designed in the post-editing and annotation stage. As a result, it is convenient for researchers to use them for testing NLP models, choosing any subset portion or combination, or cross validation usage.

Error Examples from English Corpus Fixed in AlphaMWE

Some error annotations of vMWEs in source monolingual corpus surly have some impact on the accuracy level of the *vMWE discovery and identification* shared task, but also affect the bilingual usage of AlphaMWE, so we tried to address all these cases. For instance, the example sentence in Fig. 5, English corpus annotated wrongly the sequence "had to go on" as a verbal idioms (VIDs) which is not accurate. The verb "had" here is affiliated with "all he had" instead of "to go on". So either we shall annotate "go on" as vMWE in the sentence or the overall clause "*all he had to go on*" as a studied term.

Another example with a different type of vMWE is the sentence "He put them on in a kind of trance." where the source English corpus tagged "put" and "trance" as Light-verb construction

Plain English Corpus	The chair was comfortable, and the beer had gone slightly to his head. I was smoking my pipe quietly by my dismantled steamer, and saw them all cutting capers in the light, with their arms lifted high, when the stout man with mustaches came *tearing down* to the river, a tin pail in his hand, assured me that everybody was 'behaving splendidly, splendidly, dipped about a quart of water and *tore back* again. (*the italic was not annotated in source English*)
English MWEs	gone (slightly) to his head, cutting capers, tearing down, tore back
Target Chiense Corpus	椅子很舒服，啤酒已经微微让他上了头。[sourceVMWE: gone (slightly) to his head][targetVMWE: (微微)让他上了头] 我在拆开的汽船旁静静地抽着烟斗，看到他们都在灯光下欢呼雀跃，高举双臂，这时，那个留着胡子的大块头，手里拿着一个铁皮桶，快速来到河边，向我确保大家都"表现得很精彩，很精彩"，他浸了大约一夸脱的水，又快速回去了。[sourceVMWE: cutting capers; tearing down; tore back][targetVMWE: 欢呼雀跃; 快速到; 快速回去]
Target German Corpus	Der Stuhl war bequem, und das Bier war ihm leicht zu Kopf gestiegen. [sourceVMWE: gone (slightly) to his head][targetVMWE: (leicht) zu Kopf gestiegen] Ich rauchte leise meine Pfeife an meinem zerlegten Dampfer und sah, wie sie alle im Licht mit hoch erhobenen Armen Luftsprünge machten, als der stämmige Mann mit Schnurrbart mit einem Blecheimer in der Hand zum Fluss hinunterkam und mir versicherte, dass sich alle "prächtig, prächtig benahmen, etwa einen Liter Wasser eintauchte und wieder zurückwankte". [sourceVMWE: cutting capers; tearing down; tore back][targetVMWE: Luftsprünge machten; hinunterkam; zurückwankte]
Target Polish Corpus	Krzesło było wygodne, a piwo lekko uderzyło mu do głowy. [sourceVMWE: gone (slightly) to his head][targetVMWE: (lekko) uderzyło mu do głowy] Cicho paliłem swoją fajkę przy zdemontowanym parowcu i widziałem, jak wszyscy pląsają w świetle, z podniesionymi wysoko ramionami, gdy twardziel z wąsami przyszedł szybkim krokiem do rzeki, blaszany wiaderko w dłoni, zapewnił mnie, że wszyscy "zachowują się wspaniale, wspaniale, nabrał około ćwiartkę wody i zawrócił szybkim krokiem". [sourceVMWE: cutting capers; tearing down; tore back][targetVMWE: pląsają; przyszedł szybkim krokiem; zawrócił szybkim krokiem]
AlphaMWE corpora examples from multilingual parallel files. "cutting capers" was annotated as VID type of MWEs, while "tearing down" and "tore back" were not annotated in the source English corpus. We added them into AlphaMWE multilingual corpora since they do cause translation errors for most state-of-the-art MT models. The bilingual MWEs are aligned with their appearance order from sentence inside the afterwards attached bracket-pairs.	

Figure 11: AlphaMWE corpora samples with two sentences

(VLC.cause). However, the phrase is with "put...on" instead of "put...trance". "put someone into a trance" is a phrase to express "make someone into a half-conscious state". However, for this sentence, if we check back a bit further of the context, it means "he put on his cloth in a kind of trance". The word "trance" is affiliated with the phrase "*in a kind of trance*" instead of "put".

Appendix B: English→German/Polish MT Samples Reflecting Afore Mentioned MWE Related Issues.

English→German

Firstly, for the English vMWE translates into single German word, let's see the vMWE "woke up" the sentence "An old woman with crinkly grey hair woke up at her post outside the lavatory and opened the door, smiling and grasping a filthy cleaning rag." has corresponding German aligned word "erwachte" with a suitable translation "Eine alte Frau mit krausem, grauem Haar erwachte auf ihrem Posten vor der Toilette und öffnete die Tür, lächelte und griff nach einem schmutzigen Putzlappen.".

This also occurs in English to Chinese translation, such as an English verb+particle MWE getting aligned to one single Chinese character/word. For example, in this sentence "The fact that my name has been mixed up in this.", the vMWE (VPC) *mixed up* gets aligned to single character word "混(hùn)" in a suitable translation "事实上，我的名字已经被混在这里面了。(Shì shí shàng, wǒ de míng zì yǐ jīng bèi hùn zài zhè lǐ miàn le)".

Secondly, for the automatic translation to German that is very *biased* towards choosing the

polite or formal form, see the examples such as "Sie"instead of the second form singular "du"for "you", "auf Basis von" instead of "basierend auf" for "based on". To achieve a higher accuracy level of MT, it shall depend on the context of usage to decide which form is more suitable.

Thirdly, for the English verbal multiword expressions that are often not translated as verbal multiword expressions to German. This indicates some further work to explore by MT researchers to develop better models to have the machine producing corresponding German existing MWEs.

English→Polish

Regarding the MT output issues on English to Polish that fall into coherence-unaware error, for instance, the vMWE "write off" in sentence "Then someone says that it can't be long now before the Russians write Arafat off." was translated as "Wypiszą" (Potem ktoś mówi, że już niedługo Rosjanie wypiszą Arafata.) which means "prescribe", instead of correct one "spiszą na straty (Arafata)". This error shall be able to avoid by the coherence of the sentence itself in meaning preservation models.

For the literal translation, we can see the example vMWE "gave (him) a look" in the sentence "She ruffled her feathers and gave him a look of deep disgust." which was literally translated as "dała mu spojrzenie", however, in Polish, people use "**throw** a look" as "rzuciła (mu) spojrzenie" instead of "gave (dała, a female form)"[4]. Another example of literal translation leading to errors is the vMWE "turn the tables" from sentence "Now Iran wants to turn the tables and is inviting cartoonists to do their best by depicting the Holocaust." which is translated as "odwrócić stoliki (turn tables)", however, it shall be "odwrócić sytuację (turn the situation)" or "odwrócić rolę (turn role)" with a proper translation "*Teraz Iran chce odwrócić sytuację i zachęca rysowników, by zrobili wszystko, co w ich mocy, przedstawiając Holocaust.*" These two examples present the localization issue in the target language.

For the context unaware issue, we can look back to the example sentence "But it did not give me the time of day." from Fig. 8. It was literally translated word by word into "Ale nie dało mi to pory dnia." which is in the sense of hour/time. However, it shall be "Nie sądzę aby to było coś wyjątkowo/szczególnie dla mnie. (I do not think this is special to me.)" based on the context, or "Ale to nie moja bajka" as an idiomatic expression which means "not my fairy tale" (indicating *not my cup of tea*).

Appendix C: Initial Post-editing and Annotation Contact List (Fig.12).

en->zh	Lifeng Han, <lifeng.han@adaptcentre.ie> ADAPT Research Centre, DCU, Dublin, Ireland
	Ning Jiang, <njiang@tcd.ie> School of Linguistic, Speech and Communication Sciences, TCD, Ireland
	Qinyuan Li, <liq3@tcd.ie> School of Education, Trinity College Dublin (TCD), Ireland
	Pan Pan, <panpan@m.scnu.edu.cn> School of Foreign Studies, South China Normal University, Guangzhou, China
en->de	Gültekin Cakir, <gueltekin.cakir@mu.ie> Innovation Value Institute, Maynooth University, Ireland
	Daniela Gierschek, <daniela.gierschek@uni.lu> Institute of Luxembourgish Linguistics and Literature, Université du Luxembourg, 2 Avenue de l'Université, 4365 Esch-sur-Alzette, Luxembourg
	Vanessa Smolik, <v.smolik@uni-bielefeld.de> Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany
en->pl	Teresa Flera, <t.flera@uw.edu.pl> Doctoral School of Humanities (Institute of English Studies), University of Warsaw, Poland
	Sonia Ramotowska, <s.ramotowska@uva.nl> Institute for Logic, Language and Computation, University of Amsterdam, Science Park 107, 1098 XG Amsterdam, The Netherlands

Figure 12: AlphaMWE corpora initial contact list (with alphabetical order)

[4]a proper translation: *Nastroszyła sobie pióra i rzuciła mu spojrzenie głębokiego obrzydzenia.* Also the MT output word for "Nastroszyła" was "Zdruzgotała" which is wrong meaning.