

Multi-word Expressions in English-Latvian Machine Translation

Inguna SKADIŅA

University of Latvia, Institute of Mathematics and Computer Science
Raiņa bulv. 29, Riga, Latvia

`inguna.skadina@lumii.lv`

Abstract. The paper presents series of experiments that aim to find best method how to treat multi-word expressions (MWE) in machine translation task. Methods have been investigated in a framework of statistical machine translation (SMT) for translation from English into Latvian. MWE candidates have been extracted using pattern-based and statistical approaches. Different techniques for MWE integration into SMT system are analysed. The best result - +0.59 BLEU points – has been achieved by combining two phrase tables - bilingual MWE dictionary and phrase table created from the parallel corpus in which statistically extracted MWE candidates are treated as single tokens. Using only bilingual dictionary as additional source of information the best result (+0.36 BLEU points) is obtained by combining two phrase tables. In case of statistically obtained MWE lists, the best result (+0.51 BLEU points) is achieved with the largest list of MWE candidates.

Keywords: multi-word expressions, statistical machine translation, bilingual dictionary, under-resourced languages, Latvian language

1. Introduction

Multi-word expressions (MWEs), which often are defined as “lexical items that (a) can be decomposed in multiple lexemes and (b) display lexical, syntactic, semantic, pragmatic and/or statistical idiomaticity” (Baldwin and Kim, 2010), is well known ‘pain in a neck’ (Sag et al., 2002) for human language technology researchers.

MWEs form substantial part of our language. According to Sag, 41% of entries in WordNet (Fellbaum, 2005) are MWEs. The best known type of multi-word expressions are idioms (e.g. ‘feeling blue’, ‘kick the bucket’, etc.). However, idioms are only one group of MWEs; many terms (e.g. ‘frog brick’, ‘tablet computer’), phrasal verbs (e.g. ‘give up’), named entities (e.g. ‘president Bush’, ‘New York’), nominal phrases (e.g. ‘traffic light’) etc. are also classified as MWEs.

MWEs and specific classes of MWEs have been researched for many years. However, lot of problems are still left unsolved (Savary et al., 2015). One of them is MWE processing in automatic translation task. Using rule-based approach MWEs need to be encoded into dictionary. MWEs that are not included into dictionary usually are treated as common phrases and often are translated incorrectly. In case of statistical machine translation (SMT) MWEs need to be learned from the text corpora during a training process. Common and frequently used phrases, including MWEs, usually are

recognized during training process and are stored into phrase table. However, less frequent and longer phrases are usually missing. In many cases this leads to wrong, word by word translation of MWEs.

For MWE translation different approaches are described in literature and has been applied for different language pairs. However, most of existing studies concentrate on widely used languages, such as English, French (e.g., Bouamor et al., 2012) or Spanish (e.g., Lambert and Banchs, 2006). Moreover, most of these studies deal with languages with rather simple morphology and fixed word order.

The aim of this paper is to investigate possible ways of MWE treatment in English-Latvian automated translation task in a framework of statistical machine translation. The paper continues research presented in Skadiņa (2016) through systemic and comprehensive analysis and comparison of different approaches for MWE identification, alignment and treatment in automated translation task. It is first such study for the Latvian language and could be useful for other morphologically rich under-resourced languages. The obtained evaluation results show improvement in automatic evaluation by the BLEU score, as well as increase of fluency and adequacy of translation.

2. Related work

Translation of MWEs has been recognized as a problem already at rather early stage of machine translation (Hutchins and Somers, 1992). During decades when rule-based systems were dominant, special MWE dictionaries were created manually or semi-automatically. Such approach was also chosen by Deksnis et al. (2008) for English-Latvian rule-based machine translation system. For MWE treatment authors proposed to use special dictionary of MWEs and include additional MWE processing step during parsing.

During last decade statistical machine translation has become dominant and thus ways how to teach SMT systems translate MWEs correctly have attracted attention of researchers. Two actual issues have been studied – how to create/obtain MWE dictionary, if such dictionary does not exist? How to integrate MWE dictionary into SMT system?

For MWE identification two main techniques are pattern-based approach (e.g. Ramisch, 2015; Pinnis et al., 2012) and statistical approach (Pecina, 2005). More complicated issue is MWE alignment between languages. Here mostly dictionary-based, context-based and statistical approaches are applied (e.g., Ștefănescu, 2012, Macken et al., 2008).

Different techniques how to integrate MWE dictionary into SMT system at first were investigated by Carpuat and Diab (2010). They analyse two approaches – static (MWEs are treated as single units) and dynamic (special feature is used to indicate presence of MWE) for English-Arabic SMT. The best result has been achieved with static approach. It needs to be mentioned that MWEs in this study were obtained from WordNet. Two years later Bouamor et al. (2012) investigated three strategies for integration of bilingual MWE dictionary in English-French SMT task - retraining with MWEs as parallel corpus, MWEs in a phrase table and a specific feature in phrase table. The best results were obtained with the retraining approach.

For less resourced languages several classes of MWEs have been investigated in context of SMT. Pinnis and Skadiņš (2012) investigated term translation problem for

domain specific SMT. They reported transformation of translation model into term-aware phrase tables using specific feature as the most successful approach. Following recommendations by Carpuat and Diab (2010), Kordoni and Simova (2014) analysed phrasal verb translation with help of dictionary in English-Bulgarian SMT task. They report dynamic integration (specific feature) as best approach. Recently Cholakov and Kordoni (2016) applied word embeddings to augment the phrase table of an English-Bulgarian SMT system with new features. This approach outperformed their previous results for SMT of phrasal verbs.

3. Data and tools

The DGT-TM corpus (Steinberger et al., 2012) of legal documents was used in described experiments. Although the corpus does not contain idiomatic expressions, it contains a lot of terminological units, light verb constructions and named entities, that needs to be treated as MWEs. The training corpus contains 1,63 million unique English-Latvian parallel sentences. Tuning and test data were selected randomly and separated from the training data before experiments started. 1000 sentences were used as test data and 2000 sentences were used as tuning data. For corpus cleaning and selection of test and tuning data LetsMT! (Vasiljevs et al., 2012) platform was used.

The *Moses* toolkit (Keohn et al., 2007) with default settings was used for training and translation. The 5-gram language model was created with KenLM (Heafield et al., 2013), minimum error-rate training (Och, 2003) was used for tuning and BLEU score (Papineni et al., 2002) was used for automatic evaluation.

For morphological tagging state of art taggers - TreeTagger (Schmid., 1995) for English and Latvian tagger by Paikens (2013) – were used.

4. Two approaches for MWE identification

4.1. Pattern-based approach

Identification of MWE candidates

One way how to identify MWEs in a text is to apply morpho-syntactic patterns that extract all phrases which match particular pattern. Usually this approach leads to overgeneration, thus to filter reliable MWE candidates the association measures are applied afterwards.

214 patterns for Latvian and 61 pattern for English were created for MWE identification. Most of these patterns describe noun phrases. Latvian language is inflected language, thus more patterns, to describe correctly MWE candidates and to avoid unnecessary overgeneration, are necessary.

Fig 1 illustrates typical pattern for Latvian noun phrase identification.

To find MWE candidates and mark them in corpora the freely available *mwetoolkit* (Ramisch, 2015) was used. The toolkit allows to identify and extract list of MWE candidates using user defined morphosyntactic patterns. It also calculates different association measures - Maximum Likelihood Estimator, Pointwise Mutual Information, Student's t test score, Dice's Coefficient, and Log-likelihood. Finally, it allows to mark MWE candidates in text.

```
<pat>
  <w pos="n*g*" /> <w pos="affpn*" /> <w pos="ncfpn*" />
</pat>
```

Fig 1. Example of pattern that identifies noun phrase consisting of noun (n) in genitive (g) followed by plural (p) feminine (f) adjective (a) in nominative case (n) and governed by feminine (f) plural (p) noun (n) in nominative case (n).

Using *mwetoolkit* 610 thousand unique MWE candidates for English and 3.68 million candidates have identified. Such a big difference in number of extracted candidates can be explained by rich morphology of the Latvian language and overgeneration. The MWE candidates then were marked in a text ¹.

Creation of bilingual dictionary

When MWE candidates are identified in a monolingual text, a bilingual MWE dictionary could be built through alignment of possible translation equivalents. To find translation equivalents for monolingual MWE candidates MPAligner toolkit (Pinnis, 2013) was applied. The toolkit initially was designed to find translation equivalents for terminological units (single token as well as multiword), however, it can be applied for other alignment tasks as well. MPAligner can be used in two ways - with and without dictionary. In a first case translation equivalents will be identified using transliteration, in second case both transliteration and dictionary-based alignment will be performed.

```
activity breakdown ||| darbības iedalījuma ||| 0.814912
activity category ||| darbības kategorija ||| 0.738657
activity class ||| darbības klasi ||| 0.777126
activity code ||| darbības kodam ||| 0.774599
activity code ||| darbības kods ||| 0.909243
activity code ||| darbības kodu ||| 0.872991
activity compensation ||| darbības kompensācijai ||| 0.831963
activity concentration ||| darbības koncentrācijas ||| 0.888705
activity control ||| darbības kontrole ||| 0.860559
activity description ||| darbības aprakstā ||| 0.799606
activity field ||| darbības joma ||| 0.831887
```

Fig 2. Fragment of extracted bilingual dictionary with reliability scores.

In our experiments we used MPAligner with the dictionary that is included in the toolkit. The tool at first extracts all possible translations of MWEs and then selects those that are above specified threshold. For our experiments we used the default threshold 0.7. Initially the toolkit extracted 369 506 candidate pairs (including duplicates). After filtering, 55 363 pairs were kept.

¹ In first experiments Dice's coefficient was applied afterwards to keep most reliable MWE candidates. However, it was discovered, that unreliable candidates are efficiently filtered during alignment step.

Fig 2 shows fragment of the extracted dictionary and alignment probabilities. One could notice that for single English phrase several Latvian phrases are identified which correspond to different inflectional forms of the phrase.

4.2. Association measures for MWE identification

Another way for MWE identification is application of different association measures. Such approach allows to identify MWEs that are hard to describe by patterns. However, this approach also recognizes different frequently used strings as MWE candidates that are not grammatically correct phrases (e.g. 'of the').

The *Collocate* tool version 1.0, which "is designed to provide information about the collocations in a text or corpus" (Barlow, 2004) was used for MWE candidate extraction. The tool allows to find collocations using different association measures: mutual information, T-score, and Log Likelihood. At first several experiments were performed with different association measures to select the most appropriate. After manual inspection of most frequent collocates, it was decided to use the Log Likelihood for collocation extraction.

Several thresholds have been used to extract MWE candidates. The extracted MWE candidates then were filtered using regular expressions to exclude ungrammatical phrases, e.g. preposition followed by determiner or phrases with numbers, etc. In addition, top 200 phrases were manually checked. Table 1 summarizes statistics about MWE candidates after application of different filters.

Table 1. Number of extracted MWE candidates using different filters and thresholds

Method	Minimal frequency	Minimal log-likelihood	Number of English MWE candidates	Number of Latvian MWE candidates
Data extracted with <i>Collocate</i> tool	2	3	1,707,306	1,087,932
Filtering with regular expressions	2	3	1,155,360	795,063
Filtered with regular expressions	3	10	1,074,113	556,695
Data extracted with <i>Collocate</i> tool	10	10	442,898	565,087
Filtered with regular expressions	10	10	98,843	
Data extracted with <i>Collocate</i> tool	5	10		598,868
Filtered with regular expressions	5	10		88,943

5. Strategies for MWE integration into SMT system

Using extracted MWE candidates several strategies how MWEs could be integrated into SMT system have been investigated.

5.1. Integration of bilingual MWE dictionary into SMT system

Three different strategies how bilingual MWE dictionary could be integrated into SMT system were investigated:

1. MWE dictionary data were added to the training data and an SMT system was retrained.
2. Additional phrase table was created from bilingual dictionary. Scores assigned by MPAligner were used as translation probabilities for the second phrase table. Translation tables were then combined in two different ways – by scoring with both tables and scoring with either table.
3. Additional feature that indicates presence of MWE was added to the phrase table. The phrase table combination technique proposed by Bisazza et al (2011) was applied to create MWE aware translation table.

All three approaches were evaluated with BLEU metrics. Obtained results are summarised in Table 2.

Table 2. Application of different strategies for integration of bilingual MWE dictionary into SMT system

Method	BLEU
Baseline	46.35
D1-1: Adding MWEs to training data	46.49
D1-2: Adding MWEs to training data (larger dictionary)	46.39
D2-1: Two translation tables (scoring with both tables)	46.46
D2-2: Two translation tables (scoring with either table)	46.71
D3: Additional feature	46.40

Automatic evaluation results for three initial systems (A1-1, A2-1 and A3) were close to the baseline. Thus two additional experiments were performed. In first experiment the alignment threshold of MPAligner tool was lowered to 0.4, allowing to create larger bilingual dictionary (83,295 entries). This experiment (A1-2) did not lead to significant improvement of BLEU score.

In the second experiment (A2-2) two translation tables were scored with either table. With this approach the best result - +0.36 BLEU point improvement – was achieved. Obtained results differ from results presented by Pinnis and Skadiņš (2012) who reported additional feature as most efficient way for terminology treatment. This could be explained by quality of the baseline system – while BLEU scores for baseline systems in Pinnis and Skadiņš did not exceed 16 BLEU points, baseline system presented in this paper received 46.35 BLEU points.

A manual investigation which was performed to find main differences in translation, revealed some improvement in fluency and adequacy of translations (Skadiņa, 2016). However, it also revealed some limitations. One limitation is related to creation of bilingual MWE dictionary – it is limited to defined patterns, dictionary used for alignment and threshold applied for dictionary filtering.

5.2. Integration of MWE lists into SMT system

As integration of bilingual MWE dictionary in most of cases demonstrated small improvement in terms of BLEU score, another approach – integration of monolingual MWE lists, which were extracted with association measures, into SMT system - was investigated. Decision to integrate monolingual lists instead of bilingual dictionary was also chosen after manual inspection of top 200 MWE candidates in each language, which showed that these lists contain different MWEs. This approach also helped to solve one of deficiencies of the previous approach – lack of MWEs that are translated as single token. Table 3 summarizes automatic evaluation results for systems created from different MWE candidate lists.

Table 3. Results of automatic evaluation

System	BLEU
Baseline	46.35
S1: Minimal frequency >3	46.86
S2: Frequency>2 and cost >9	44.57
S3: Frequency for Latvian >4, freq. for English >9;cost>9	45.13

The best result - +0.5 BLEU point improvement - has been achieved with rather noisy data, while results for other systems with less data did not exceed the baseline. Similarly to the previous experiment, manual investigation of translations have revealed some improvements regarding adequacy and fluency of obtained translation.

5.3. Hybrids: combining data and approaches

Several experiments that combine MWEs obtained with the methods described above have been performed to find best combination of knowledge-base and statistical approaches.

The first experiment was influenced by the best obtained result so far – treatment of MWEs as single unit during training. This approach (C1) has been applied on data from bilingual dictionary, resulting in better result (46.68 BLEU; +0.33), when compared with most of previous results, obtained with bilingual dictionary.

In next experiment (C2) both data sources - MWEs from the bilingual dictionary and the largest MWE lists - were merged and marked (concatenated) in training data. This approach lead to slightly better results in terms of the BLEU score (46.72; +0.37), but did not exceed simple concatenation method presented in previous chapter.

Finally, four experiments were performed by combining two phrase tables from previous experiments. Systems C3 and C4 combine the phrase table obtained by adding bilingual dictionary to the training data (system D1) and phrase table obtained by treatment of MWEs from the largest MWE list as single units (system S1). Systems C5 and C6 combine the phrase table created from bilingual dictionary and phrase table obtained by treatment of MWEs from the largest MWE list as single units (system S1). The best result (+0.59 BLEU points) was achieved by system C6, which uses phrase table created from bilingual dictionary as the first.

Automatic evaluation results of all experiments are summarized in Table 4.

Table 4. Results of automatic evaluation

System	BLEU
Baseline	46.35
C1: Bilingual dictionary with concatenation approach	46.68
C2: Bilingual and monolingual MWE candidates are marked in a training data	46.72
C3: Two translation tables: system with bilingual dictionary (A1) and MWE lists (S1) (scoring with both tables)	46.42
C4: Two translation tables: system with bilingual dictionary (A1) and MWE lists (S1) (scoring with either table)	46.69
C5: Two translation tables : MWE lists (S1) and MWE dictionary (scoring with either table)	46.58
C6: Two translation tables : MWE dictionary and MWE lists (S1) (scoring with either table, bilingual dictionary first)	46.94

6. Analysis and discussion

As it has been demonstrated that increase of BLEU score does not always mean better quality of translation (Smith et al., 2016) and to analyse influence of developed methods on MWE translation, manual analysis of the output of the baseline system and the best system from each approach (D2-2, S1, and C6) was performed. 233 sentences from 1000 sentences in test corpus were translated identically by all systems, including 57 sentences which were translated identically to reference (human) translation. The baseline system generated identical translation to reference translation in 84 cases, system that uses bilingual dictionary (D2-2) generated identical translation to reference translation in 80 cases, system that uses statistically extracted MWE lists (S1) in 84 cases, but combined system in 80 cases.

For each sentence in test corpus the standard deviation was calculated between BLEU scores of these four systems. 100 sentences with higher standard deviation were analysed manually. The first observation was, that highest BLEU scores (and corresponding translations) in many cases are assigned to two systems – either baseline and D2-2 system (39 cases) or S1 system and hybrid system (20 cases). In 3 cases only the baseline system received the highest score, in 4 cases only the D2-2 system received the highest score, in 2 cases only single token approach has received the best assessment, while in 3 cases only combined system has received the highest BLEU score. Although 12 cases are insufficient to make generalization, 3 cases for each SMT system where particular system has received highest BLEU score and where the standard deviation is large, are analysed in this chapter.

Table 5 shows three cases when the baseline system has received the highest BLEU score. In all cases output from systems S1 and C6 are identical, which can be explained by architecture of C6 system.

Table 5. Examples of SMT system output when baseline system has received the highest BLEU score

Example 1		
Source	BLEU	Sentence (translation)
English source		specify the competent <u>court or arbitration tribunal</u> to hear disputes .
Reference		nosaka <u>tiesu vai šķīrējtiesu</u> , kas ir kompetenta izskatīt strīdus .
Baseline	54,91	nosaka tiesu , kas ir kompetenta izskatīt <u>strīdus vai šķīrējtiesā</u> .
D2-2	49,61	nosaka tiesu , kas ir kompetenta izskatīt strīdus <u>tiesā vai šķīrējtiesā</u> .
S1, C6	8,44	norādīt <u>kompetentā tiesa vai šķīrējtiesa</u> noteikusi izskatīt strīdus ;
Example 2		
English source		the electricity consumption measuring equipment is physically integrated in the rolling stock .
Reference		elektroenerģijas patēriņa mērīšanas iekārtas ir fiziski <u>iebūvētas</u> ritošajā sastāvā .
Baseline	65,80	elektroenerģijas patēriņa mērīšanas iekārtas ir fiziski <u>integrēta</u> ritošajā sastāvā .
D2-2	53,72	elektroenerģijas patēriņa mērīšanas iekārtas ir fiziski <u>integrēta</u> ritošā <u>sastāvā</u> .
S1, C6	26,20	elektroenerģijas patēriņa mērīšanas iekārtas , <u>kas</u> ir fiziski <u>integrēta</u> ritošo <u>sastāvu</u> .
Example 3		
English source		in taking its decision , the group supervisor shall duly consider the following :
Reference		pieņemot lēmumu , grupas uzraudzības iestāde pienācīgi ņem vērā
Baseline	88,11	pieņemot lēmumu , grupas uzraudzības iestāde pienācīgi ņem vērā :
D2-2	63,89	pieņemot lēmumu , grupas uzraudzības iestāde pienācīgi apsver šādu :
S1, C6	57,07	pieņemot lēmumu , grupas uzraudzības iestāde pienācīgi apsver šādu informāciju :

In first case, although the baseline system has been automatically evaluated as the best, the translation of the system D2-2 is more adequate and fluent, while translations from systems S1 and C6 are incomprehensible. It could be explained by more adequate translation of MWE ‘court or arbitration tribunal’, where system D2-2 kept agreement between constituents, while the baseline system has lost it. In second case the output of the baseline system is more fluent as for other systems. No influence on MWE translation observed. In third case translation of the baseline system is identical to reference translation, except column, which is missing in reference translation. Fluent, and almost adequate translation is generated also by systems S1 and C6. The translation

of the MWE ‘shall duly consider the following’ by these systems includes ‘šādu informāciju’ (following information), while reference and baseline system ignores word ‘following’ in translation.

Table 6. Examples of SMT output when dictionary based system has received the highest BLEU score

Example 1		
Source	BLEU	Sentence (translation)
English source		fish , short-term toxicity test on embryo and sac-fry stages
Reference		īstermiņa toksicitātes tests zivīm embrija un dzeltenummaisā attīstības posmos
Baseline	8,05	īslaicīgas iedarbības toksicitātes tests zivju embriju un pieņu / ikru attīstības stadijas
D2-2	88,01	par īstermiņa toksicitātes tests zivīm embrija un dzeltenummaisā attīstības posmos
S1, C6	5,30	zivju embriju un pieņu / ikru attīstības stadijas īstermiņa toksiskuma tests
Example 2		
English source		fats , japan wax , rosin and soaps
Reference		tauki , japānas vasks , kolofonijš un ziepes
Baseline	19,13	tauki , <u>sumahu</u> vasks , kolofonija un ziepes
D2-2	59,46	tauki , japānas vasks , kolofonija un ziepes
S1, C6	14,53	tauki , vasks ; <u>sumahu</u> vasks , kolofonija un ziepes
Example 3		
English source		separately , the cif price , <u>per tonne of product</u> , in euros , <u>corresponding to the tender</u> .
Reference		atsevišķi , cif cenu <u>par tonnu produkta</u> euro <u>atbilstoši iesniegtajam projekta pieteikumam</u> .
Baseline	19,10	atsevišķi , cif cena <u>par vienu tonnu produkta</u> euro , <u>kas atbilst konkursa</u> .
D2-2	57,60	atsevišķi , cif cenu <u>par tonnu produkta</u> euro , <u>kas atbilst piedāvājumu</u> .
S1, C6	12,45	atsevišķi , ir paredzēts , ka cif cenas tonnu produkta euro , <u>kas atbilst piedāvājumu</u> .

Table 6 presents cases when dictionary-based approach outperforms other approaches in terms of BLEU score. All three cases clearly demonstrate positive influence of bilingual MWE dictionary. In first example two MWEs ‘short term toxicity test’ (īstermiņa toksicitātes tests) and ‘sac-fry stages’ (dzeltenummaisā attīstības posmos) is correctly translated only by system D2-2. It also needs to be mentioned, that output from systems S1 and C6 is more adequate and fluent when compared to baseline. Similarly, in second example the term ‘japan wax’ (japānas vasks) is correctly translated

only by system D2-2. Finally in third example, MWE ‘per tonne of product’ is correctly translated by system D2-2 (the translation of the baseline system is also adequate). However, none of the systems provide completely correct translation for ‘corresponding to the tender’, while output of improved system generated incorrect inflection.

Table 7 summarizes cases when MWE lists treated as single tokens outperform other approaches in terms of BLEU score. In first case output of system S1 is identical reference translation, as it could be interpreted as MWE. However, translations of baseline and D2-2 systems are also fluent and rather adequate, while output from system C6 is not adequate, although it is fluent and is evaluated as the second best translation. In second case, systems S1 and C6 translates MWE ‘draw up’ identically to reference translation, but fails in translation of ‘analytical accounts’. The most adequate and fluent translation is generated by D2-2 system. The third example illustrates translation of two MWEs ‘set out’ and ‘the requirement concerning professional secrecy’ which translated identically to reference translation by systems S1 and C6. However, these translations include additional word ‘uz’ (on). Thus translations of baseline and D2-2 systems are more fluent.

Table 7. Examples of SMT output when MWE lists has received the highest BLEU score

Example 1		
Source	BLEU	Sentence (translation)
English source		legislative procedure ongoing
Reference, S1	100,00	notiek likumdošanas procedūra
Baseline, D2-2	37,99	uzsākta leģislatīvā procedūra
C6	42,89	leģislatīvā procedūra
Example 2		
English source		the agency shall draw up analytical accounts of its revenue and expenditure .
Reference		aģentūra veic analītisku ieņēmumu un izdevumu uzskaiti .
Baseline	5.93	aģentūra izstrādā analītisko uzskaiti par tā ieņēmumiem un izdevumiem .
D2-2	6.74	aģentūra izstrādā analītisko uzskaiti par ieņēmumiem un izdevumiem .
S1, C6	70,71	aģentūra veic analītisku ieņēmumu un izdevumu tāmi .
Example 3		
English source		such information is subject to the requirement concerning professional secrecy set out in article 37 of the statute of the escb .
Reference		uz šādu informāciju attiecas dienesta noslēpuma pienākums saskaņā ar ecbs statūtu 37 . pantu .
Baseline, D2-2	31.47	uz šādu informāciju attiecas dienesta noslēpuma prasības , kas attiecas uz 37 . pantā izklāstītajiem ecbs statūtiem .
S1, C6	87.61	uz šādu informāciju attiecas dienesta noslēpuma pienākums saskaņā ar ecbs statūtu 37 . pantu uz .

Table 8 summarizes cases when combination of both approaches outperforms each separate method in terms of BLEU score.

Table 8. Examples of SMT output when hybrid approach has received the highest BLEU score

Example 1		
Source	BLEU	Sentence (translation)
English source		the council also notes that coordination arrangements are required in tbilisi as well as in brussels .
Reference		padome arī ņem vērā to , ka koordinēšanas pasākumi vajadzīgi gan tbilisi , gan briselē .
Baseline	30,33	padome arī norāda , ka koordinācijas pasākumi būtu jāveic tbilisi , gan briselē .
D2-2	10,13	padome arī norāda , ka koordinācijas pasākumi būtu jāveic tbilisi , kā arī briselē .
S1	9,44	padome arī norāda , ka koordinācijas pasākumi būtu jāveic gan briselē vadītājam tbilisi .
C6	44,29	padome arī ņem vērā to , ka koordinācijas pasākumi būtu jāveic gan briselē vadītājam tbilisi .
Example 2		
English source		to enable the commission to plan <u>its safeguards activities</u> , the persons or undertakings referred to in the first subparagraph of article 3(1) shall communicate to the commission the following information :
Reference		lai dotu iespēju komisijai plānot <u>drošības pasākumu darbības</u> , personas vai uzņēmumi , kas minēti 3 . panta 1 . punkta pirmajā daļā , paziņo komisijai šādu informāciju :
Baseline, D2-2	49,96	lai komisija varētu plānot <u>savu drošības pasākumos</u> , personas vai uzņēmumi , kas minēti 3 . panta 1 . punkta pirmo daļu komisijai paziņo šādu informāciju :
S1	74,92	lai komisija varētu plānot <u>savas drošības pasākumu darbību</u> , personas vai uzņēmumi , kas minēti 3 . panta 1 . punkta pirmajā daļā , paziņo komisijai šādu informāciju :
C6	75,65	lai komisija varētu plānot <u>savas darbības drošības pasākumu</u> , personas vai uzņēmumi , kas minēti 3 . panta 1 . punkta pirmajā daļā , paziņo komisijai šādu informāciju :

Example 3		
English source		establishing a prohibition of fishing for plaice <u>in areas viif and viig</u> by vessels flying the flag of the united kingdom
Reference		ar ko nosaka aizliegumu apvienotās karalistes karoga kuģiem zvejot jūras zeltpleksti <u>viif un viig zonā</u>
Baseline, D2-2	21.63	ar ko nosaka aizliegumu kuģiem zvejot jūras zeltplekstes <u>viif un viig</u> , ko veic kuģi , kas peld ar apvienotās karalistes karogu
S1	31.67	ar ko nosaka aizliegumu kuģiem zvejot jūras zeltplekstes <u>viif un viig zonā</u> , kuri peld ar apvienotās karalistes karogu ,
C6	44.23	ar ko nosaka aizliegumu zvejot jūras zeltpleksti <u>viif un viig</u> kuģiem , kas peld ar apvienotās karalistes karogu

It needs to be mentioned that the best BLEU score for system combination approach usually is assigned to longer sentences. However, as it is illustrated in examples it does not lead to the better quality. In the first example MWE ‘notes that’ is translated by system C6 identically as in reference translation. It allows the system receive highest BLEU score, however MWE ‘as well as’ is correctly translated only by system D2-2, making this translation most fluent and adequate. In second example translations from systems S1 and C6 are close in terms of BLEU score. The most complicated here is translation of MWE ‘safeguard activities’, the baseline and D2-2 systems have lost translation for word ‘activities’, while system C6 made mistake in word ordering, and only system S1 translated it correctly, but in wrong inflectional form. The same situation can be observed in third example, in which only system S1 does not loose word ‘zona’ (area), when translating phrase ‘in areas viif and viig’.

7. Conclusion

In this paper comprehensive analysis of different methods for MWE treatment in statistical machine translation is presented. Different approaches – dynamic, static and hybrid – are analysed. The best result (+0.59 BLEU points) is achieved by hybrid approach. The manual analysis of obtained results demonstrated improvement in fluency and adequacy when MWE – aware systems were involved. However, BLEU scores assigned to translation did not always highlight the best translation. Positive influence of bilingual MWE dictionary on translation quality was observed in all analysed cases. However, it was not a case of hybrid system. Where it concerns treatment of MWEs as single token, it improves MWE translation, but not always the overall translation. Although methods are analysed for English-Latvian translation they could be applied for other languages with limited language resources as well.

Acknowledgements

The research was supported by Grant 271/2012 from the Latvian Council of Science. This work has been supported by the IC1207 COST Action PARSEME as part of its scientific program. I also would like to thank Mārcis Pinnis, Matīss Rikters and Raivis Skadiņš for their valuable comments and support.

References

- Baldwin, T., S.N. Kim, S.N. (2010). Multiword Expressions. Handbook of Natural Language Processing.
- Barlow. M. (2004). Collocate 1.0: Locating collocations and terminology.
- Bisazza, A., N. Ruiz, N., Federico, M. (2011). Fill-up versus Interpolation Methods for Phrase-based SMT Adaptation. International Workshop on Spoken Language Translation (IWSLT), 136-143.
- Bouamor, D., Semmar, N., Zweigenbaum, P. (2012). Identifying bilingual multi-word expressions for statistical machine translation. Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation, 674-679.
- Cholakov, K., Kordoni, V. (2016). Using Word Embeddings for Improving Statistical Machine Translation of Phrasal Verbs. Proceedings of the 12th Workshop on Multiword Expressions, 56-60.
- Deksne, D., Skadins, R., Skadina, I. (2008). Dictionary of Multiword Expressions for Translation into Highly Inflected Languages. Proceedings of the International Conference on Language Resources and Evaluation LREC 2008, 1401-1405.
- Fellbaum, Ch. (2005). WordNet and wordnets. Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, 665-670.
- Heafield, K., Pouzyrevsky, I., Clark, J.H., Koehn, P. (2013). Scalable Modified Kneser-Ney Language Model Estimation. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 690-696.
- Hutchins, W. J., Somers, H. (1992). An introduction to machine translation. Academic Press, London.
- Koehn, P., Hoang, H., Birch, A., Callison-Burch, Ch., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, Ch., Zens, R., Dyer, Ch., Bojar, A., Constantin, A., Herbst, E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation, Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, 177-180.
- Kordoni, V., Simova, I. (2014). Multiword Expressions in Machine Translation. Proceedings of the LREC 2014, Ninth International Conference on Language Resources and Evaluation, 1208-1211.
- Lambert, P., Banchs, R. (2006). Grouping multi-word expressions according to Part-Of-Speech in statistical machine translation. Proceedings of the EACL Workshop on Multi-word expressions in a multilingual context, 9-16.
- Och, F. (2003). Minimum Error Rate Training in Statistical Machine Translation. ACL '03 Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, 160-167.
- Paikens, P. (2013). Morphological analysis with limited resources: Latvian example. Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013), 267-277.
- Papineni, K., Roukos, S., Ward, T., Zhu, W. (2002). BLEU: a method for automatic evaluation of machine translation. ACL-2002: 40th Annual meeting of the Association for Computational Linguistics, 311-318.
- Pecina, P. (2005). An Extensive Empirical Study of Collocation Extraction Methods. Proceedings of the ACL Student Research Workshop, 13-18.

- Pinnis, M. (2013). Context Independent Term Mapper for European Languages. *Recent Advances in Natural Language Processing*, 562-570.
- Pinnis, M., Ljubešić, N., Ștefănescu, D., Skadiņa, I., Tadić, M., Gornostay, T. (2012). Term Extraction, Tagging and Mapping Tools for Under-Resourced Languages. *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012)*.
- Pinnis, M., R. Skadiņš, R. (2012). MT Adaptation for Under-Resourced Domains – What Works and What Not. *Human Language Technologies – The Baltic Perspective. Proceedings of the Fifth International Conference Baltic HLT 2012*, 176-184.
- Ramisch, C. (2015). *Multiword Expressions Acquisition: A Generic and Open Framework*. Theory and Applications of Natural Language Processing series XIV, Springer.
- Sag, I.A., Baldwin, T., Bond, F., Copestake, A., Flickinger, d. (2002). Multiword expressions: a pain in the neck for nlp. *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, CICLing'02*, 1-15.
- Savary, A., Sailer, M., Parmentier, Y., Rosner, M., Rosén, V., Przepiórkowski, A., Krstev, C., Vincze, B. Wójtowicz, G. Smørdal Losnegaard, C. Parra, E. J. Waszczuk, M. Constant, P. Osenova, V., Sangati, F. (2015). PARSEME – PARSing and Multiword Expressions within a European multilingual network. *Proceeding of the 7th Language & Technology Conference (LTC 2015)*.
- Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. 47-50.
- Skadiņa, I. (2016). Multi-word Expressions in English-Latvian SMT: Problems and Solutions. *Frontiers in Artificial Intelligence and Applications, Volume 289: Human Language Technologies – The Baltic Perspective*, 97 – 104.
- Ștefănescu, D. (2012). Mining for Term Translations in Comparable Corpora. *The 5th Workshop on Building and Using Comparable Corpora*, 98– 103.
- Aaron Smith, Christian Hardmeier, Joerg Tiedemann. (2016). Climbing Mont BLEU: The Strange World of Reachable High-BLEU Translations. *Baltic Journal of Modern Computing*, Vol. 4, 269–281.
- Steinberger, R., Eisele, A., Klocek, S., Pilos, S., Schlüter, P. (2012). DGT-TM: A freely Available Translation Memory in 22 Languages. *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012)*, 454-459.
- Tsvetkov, Y., Wintner, S. (2010). Extraction of Multi-word Expressions from Small Parallel Corpora. *Coling 2010*, 1256–1264.
- Vasiljevs, A., Skadiņš, R., Tiedemann, J. (2012). LetsMT!: A Cloud-Based Platform for Do-It-Yourself Machine Translation. *Proceedings of the ACL 2012 System Demonstrations*, 43-48.

Received September 6, 2016, accepted October 4, 2016