qtleap

quality
translation
by deep
language
engineering
approaches

Report on the state of the art of MT

DELIVERABLE D2.1

VERSION 2.5 | 2015-06-08

QTLeap

Machine translation is a computational procedure that seeks to provide the translation of utterances from one language into another language.

Research and development around this grand challenge is bringing this technology to a level of maturity that already supports useful practical solutions. It permits to get at least the gist of the utterances being translated, and even to get pretty good results for some language pairs in some focused discourse domains, helping to reduce costs and to improve productivity in international businesses.

There is nevertheless still a way to go for this technology to attain a level of maturity that permits the delivery of quality translation across the board.

The goal of the QTLeap project is to research on and deliver an articulated methodology for machine translation that explores deep language engineering approaches in view of breaking the way to translations of higher quality.

The deeper the processing of utterances the less language-specific differences remain between the representation of the meaning of a given utterance and the meaning representation of its translation. Further chances of success can thus be explored by machine translation systems that are based on deeper semantic engineering approaches.

Deep language processing has its stepping-stone in linguistically principled methods and generalizations. It has been evolving towards supporting realistic applications, namely by embedding more data based solutions, and by exploring new types of datasets recently developed, such as parallel DeepBanks.

This progress is further supported by recent advances in terms of lexical processing. These advances have been made possible by enhanced techniques for referential and conceptual ambiguity resolution, and supported also by new types of datasets recently developed as linked open data.

The project QTLeap explores novel ways for attaining machine translation of higher quality that are opened by a new generation of increasingly sophisticated semantic datasets and by recent advances in deep language processing.

www.qtleap.eu

Funded by

QTLeap is funded by the 7th Framework Programme of the European Commission.

Supported by

And supported by the participating institutions:

Faculty of Sciences, University of Lisbon

German Research Centre for Artificial Intelligence

Charles University in Prague

Bulgarian Academy of Sciences

Humboldt University of Berlin

University of Basque Country

University of Groningen

Higher Functions, Lda

Revision history

Version	Date	Authors	Organisation	Description
1.0	Dec 10, 2013	Rudolf Rosa	CUNI	First draft
1.5	Dec 13, 2013	Rudolf Rosa, Ondřej Bojar, Ondřej Dušek, Martin Popel	CUNI	Feedback from CUNI integrated
2.0	Dec 20, 2013	Rudolf Rosa, Ondřej Bojar, Ondřej Dušek, Martin Popel	CUNI	Feedback from IICT-BAS integrated
2.0	Aug 18, 2014	Gertjan van Noord	UG	agrees with sections 5, 6 and 7 (no changes made)
2.0	Sep 9, 2014	Eneko Agirre, Gorka Labaka, Nora Aranberri, António Branco	UPV/EHU, FCUL	agrees with sections 5, 6 and 7 (no changes made)
2.0	Sep 22, 2014	Kiril Simov, Petya Osenova	IICT-BAS	agrees with sections 5, 6 and 7 (no changes made)
2.0	Sep 22, 2014	Aljoscha Burchardt	DFKI	agrees with sections 5 and 6 (additions made to sections 2 and 7)
2.5	Jun 8, 2015	Martin Popel, Rudolf Rosa	CUNI	following the recommendations in the 1st year Technical Review Report and the IICT-BAS internal review

Statement of originality
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Report on the state of the art of MT

DOCUMENT QTLEAP-2015-D2.1
EC FP7 PROJECT #610516

DELIVERABLE D2.1

completion
FINAL
status
SUBMITTED
dissemination level
PUBLIC

responsible
Jan Hajič (WP2 Coordinator)
reviewers
Kiril Simov, Petya Osenova
contributing partners
CUNI

authors
Rudolf Rosa, Ondřej Bojar, Ondřej Dušek, Martin Popel

Contents

1 Overview **8**

2 Rule-based Approaches **8**

3 Syntax-uninformed SMT **9**
3.1 Word-based Statistical Machine Translation 10
3.2 Phrase-based Statistical Machine Translation 10
3.2.1 Exploiting Morphological Analysis in SMT 11
3.3 Hierarchical Models 11
3.4 Problems of Linguistically Uninformed SMT 12

4 Shallow Syntax-Based Approaches **13**
4.1 Tree-to-string Translation 13
4.2 String-to-tree Translation 14
4.3 Tree-to-tree Translation 14

5 Post-editing **14**
5.1 Human Post-editing 14
5.2 Automatic Post-editing 15
5.3 The Depfix System 15

6 Deep-syntax-based Statistical Approaches **16**
6.1 Motivation for MT based on deep syntax 16
6.2 Deep MT prototype: TectoMT 17

7 Beyond the State of the Art **18**

8 Conclusion **19**

List of Abbreviations

APE	Automatic Post-Editing
CRF	Conditional Random Fields
HMM	Hidden Markov Models
MT	Machine Translation
PB-SMT	Phrase-Based Statistical Machine Translation
RBMT	Rule-Based Machine Translation
SMT	Statistical Machine Translation
STSG	Synchronous Tree Substitution Grammar

Introduction

In this deliverable, we review the approaches employed in current state of the art machine translation systems, identifying their strengths and weaknesses. This constitutes the starting point for the QTLeap project, as we take these methods for our baseline, and base the goals of our project on surpassing their most important limitations.

1 Overview

Machine translation (MT) started with rule-based approaches. Rule-based MT in general is very difficult because reaching a broader coverage of the language needs a lot of time, manual effort and language expertise.

MT came to a revolution with the introduction of statistical machine translation (SMT), which has been the state-of-the-art in MT ever since. The early attempts at SMT were rather simplistic and assumed a word-for-word correspondence between the languages. With phrase-based SMT (PB-SMT), a big leap in quality was achieved. The inherent limitations of PB-SMT are constantly being addressed by new methods but most of the easier errors have been fixed and improving PB-SMT is getting harder and harder.

Some of the many remaining issues of PB-SMT can be tackled by syntax-based SMT, which however on average across language pairs seems to still perform worse than PB-SMT. A promising path is deep syntax-based SMT, employing deep linguistic analysis of input sentences, but there are still many obstacles on the way to its effective use and competitive performance.

2 Rule-based Approaches

Rule-based machine translation systems with handcrafted rules, such as Systran [Toma, 1977], ETAP-3 [Boguslavsky, 1995], Lucy [Alonso and Thurmair, 2003] or Apertium [Forcada et al., 2011], require experts both in linguistics and programming to build a system of a reasonable quality for a new language pair.[1] Most such systems do not exploit fully deep analysis; typically, only morphological or syntactic processing is employed and a shallow transfer is used (but see Bond et al. [2011]). ETAP-3 performs some kind of deeper processing as its transfer is inspired by the deep syntactic representation (DSyntR) and the surface syntactic representation (SSyntR) of the Meaning-Text Theory.[2] In general, existing purely rule-based systems may be useful for small closed domains or as very rough translations for under-resourced languages,[3] but they cannot reach the goals of the QTLeap project, namely producing high-quality outbound translations that ideally do not require any human post-editing. However, recent advances such as, e.g., the integration of stochastic information into the tree selection within the transfer phase [e.g. Federmann and Hunsicker, 2011b,a], have shown great potential of a hybrid combination of statistical and linguistic methods, which is the goal of QTLeap.

[1] In Apertium, it is claimed that only linguist experts with basic knowledge of Apertium are needed for a new language pair, once the engine is in place, as the engine itself is designed to be language independent.

[2] ETAP-3 is a closed-source proprietary rule-based English-Russian MT system. Therefore, the exact extent of the deep processing cannot be reliably assessed.

[3] As noted e.g. on http://wiki.apertium.org/wiki/Frequently_Asked_Questions

Currently, development and/or research on purely rule-based MT systems is extremely rare. To the best of our knowledge, Apertium is the only purely RBMT system in active development for which information about its internals is available.[4] Adding a new language pair requires a large amount of work by a language expert, thus being rather costly, especially in comparison with SMT. Thanks to a community-style approach to resource creation, Apertium currently supports 40 translation directions, with a focus on small and under-resourced languages where there is little or no competition from SMT systems. However, this approach does not scale well to new languages, as it requires Apertium-specific resources to be manually built for each individual language pair – unlike other approaches presented in the following sections, which exploit already existing general-purpose NLP tools (such as syntactic parsers), and try to avoid handcrafting of language-specific or even language-pair-specific resources and tools.[5] In recent years, several advanced features have been developed for Apertium, such as basic semantic annotation, or semantic filters based on subject fields [Duran et al., 2013].

In QTLeap, we decided to use Lucy for German, a closed-source RBMT system that has successfully been used in previous projects (Euromatrix+, QTLaunchPad), and achieves competitive results in comparison to SMT systems, successfully capturing the structural and semantic differences between German and other languages. The translation in Lucy consists of hand-written linguistic rules, which perform deep syntactic parsing of the input sentence, transfer of the resulting tree to the target language, and generation of the target sentence by employing inflection and agreement rules.

Nowadays, even when constructing a primarily rule-based system, approaches known from SMT or other statistical methods are typically employed to construct both the lexicons and the translation rules [Font Llitjós and Vogel, 2007]. Some systems with rule-based transfer exploit (statistical) language models in the synthesis phase at runtime [Habash and Dorr, 2002]. Various statistical approaches can be employed in Apertium, such as n-gram language models for paradigm selection [Sánchez-Cartagena et al., 2012], or statistical target-language models and maximum-entropy models for lexical selection [Tyers et al., 2015]. Such systems are usually denoted as *hybrid systems* and cannot be easily classified either as rule-based or statistical, as they try to get the best of both of these approaches by exploiting the differences in their strengths and weaknesses.

3 Syntax-uninformed SMT

The first SMT systems utilized virtually no linguistic information. They provided a mapping from the source to the target language that was done on the plain word forms. More recent "factored" SMT systems allow to incorporate arbitrary information for individual tokens. Including e.g. morphological tags in the model is thus easy, but any structural information remains hard to capture in a way that would lead to a significant improvement [Cettolo et al., 2008, Wang et al., 2012a,b].

[4] Several closed-source commercial systems exist that are also presumably rule-based, but very little or no information has been published about their operation, and we are thus unable to provide details about the approaches to translation they employ.

[5] Still, some resources developed for Apertium, such as morphological analyzers, taggers and generators, can be used in other applications, especially in case of under-resourced languages [Tyers et al., 2010].

3.1 Word-based Statistical Machine Translation

Statistical machine translation was introduced in the work on IBM models [Brown et al., 1988, 1990, 1993] and it was originally word-based. Although mainstream SMT moved to phrase-based models, as phrases proved to be much more useful translation units than individual words, the original IBM models (and HMM alignment [Vogel et al., 1996]) are still used as the base approach to word-alignment in tools such as GIZA++ [Och and Ney, 2000].

SMT follows the noisy channel model, according to which the probability of the translation of a sentence, i.e. the conditional probability of the target sentence given the source sentence, can be decomposed using the Bayes rule to a product of the conditional probability of the source sentence given the target sentence and the probability of the target sentence (divided by the probability of the source sentence, which is constant in the MT setting and thus can be ignored) – see Equation 1.

$$P(target|source) \propto P(source|target) \cdot P(target) \tag{1}$$

In word-based SMT, the $P(source|target)$ is modeled by the *translation model* (TM) which uses factorization into probabilities of each target word being the translation of the corresponding source word; these probabilities are estimated from word-aligned bilingual training data. $P(target)$ is modeled by a *language model* (LM), which scores the sentences generated by the TM for fluency. Usually, an n-gram language model is employed, which estimates probabilities of occurrences of word n-grams from monolingual training data.

3.2 Phrase-based Statistical Machine Translation

For most language pairs and domains, the state-of-the-art translation results are achieved by phrase-based or hierarchical phrase-based SMT systems (PB-SMT) such as Moses [Koehn et al., 2007], Joshua [Li et al., 2009], cdec [Dyer et al., 2010], or Jane [Vilar et al., 2010].

In the phrase-based approach [Och and Weber, 1998, Och et al., 1999, Koehn et al., 2003], the parallel sentences that constitute the training data are segmented into pairs of parallel *phrases* (word n-grams), which are then stored in a *phrase table* together with the frequencies of their occurrence in the training data. The translation model then uses the phrase table to predict the probability of translating a whole source phrase to a whole target phrase, rather than predicting translations of individual words.

Most current approaches are based on the method of Koehn et al. [2003], which led to the creation of Moses [Koehn et al., 2007], one of the most widely used machine translation toolkits today. Moses uses a discriminative log-linear model instead of the basic noisy channel model [Och and Ney, 2002], which enables it to combine the scores from the translation model and the language model with various features, accounting e.g. for context or grammar.

A common method for the optimization of model weights is the Minimum Error Rate Training [Och, 2003], which uses automatic translation quality metrics computed on a heldout data set to estimate the appropriateness of the parameter setting. Several alternative methods with better properties, such as higher stability, have been proposed since, from which PRO (pairwise ranking optimization) [Hopkins and May, 2011] is one of the most widely used.

3.2.1 Exploiting Morphological Analysis in SMT

The first SMT systems did not employ any analysis of the sentences apart from tokenization and case normalization (which is usually implemented simply as lowercasing all letters). While this usually works well for morphologically poor languages, such as English or Chinese, morphologically richer languages, such as Czech or Finnish, constitute a grave obstacle for such approach, as it has to face severe data sparseness problems due to the large numbers of possible word forms.

One of the successful ways of incorporating morphological analysis of the sentence are factored translation models [Koehn and Hoang, 2007]. In this approach, each token is represented by an n-tuple of attributes, e.g. ⟨word form, lemma, part-of-speech tag⟩. When applied on the *source* side, it allows to generalize over the individual word forms and use e.g. lemmas instead, although still adhering to the phrase-based approach of translating linear sequences of entities using a phrase table. When applied on the *target* side, additional language models can be used, e.g. a LM over morphological tags. Such more general LMs can be estimated more reliably given a fixed amount of data, and the same computing resources allow for longer n-grams.

When targeting a morphologically rich language, words in new forms are often needed, i.e. forms never seen in the parallel training data. Factored models in principle allow this but in practice fall into the trap of combinatorial explosion of all possible word forms. Among the possible remedies, an interesting approach is that of [Toutanova et al., 2008], who use a two-stage system. In the first stage, the authors use a PB-SMT system to translate the source sentence into a sequence of lemmas or stems. The correct inflection of each of the tokens is then generated in the second stage. The basic setup works well with rather limited parallel data [Bojar and Kos, 2010] and Fraser et al. [2012] were the first to achieve gains also with larger data.

3.3 Hierarchical Models

Hierarchical phrase-based statistical machine translation is characterized by the introduction of non-terminal symbols into the phrases. While this approach is inspired by phrase-structure syntax and the descriptions are formalized as synchronous context-free grammars, the actual structures used in hierarchical translation systems are not based on explicit linguistic annotations. Instead, they are inferred automatically from an unparsed parallel data. As a result, they are very flat and can be thought of as phrases with gaps. The introduction of hierarchical models is motivated by the inability of traditional phrase-based approaches to capture discontinuous phrases, which is believed to be one of their most significant drawbacks.

The hierarchical approach was pioneered by Chiang et al. [2005] in their Hiero system. However, this system uses chart parsing for decoding, which is incompatible with the usual left-to-right decoding approach, and therefore cannot benefit from many methods used to improve the translation quality in PB-SMT. Watanabe et al. [2006] and [Galley and Manning, 2010] address this issue by employing a left-to-right search algorithm which allows for certain kinds of discontinuous or gappy phrases and achieves performance competitive to phrase-based models.

3.4 Problems of Linguistically Uninformed SMT

Since its introduction, SMT has faced many issues, out of which many have already been successfully addressed, but many still wait to be solved either by improvements to the existing SMT systems or by designing new ways of performing machine translation.

Linguistically uninformed statistical MT systems are **limited by the absence of appropriate linguistic abstraction**.

Weaknesses of the **shallow SMT**[6] approaches are well known: problems with word order in typologically different languages, problems with capturing long-distance dependencies, grammatical and semantic cohesion, etc. [Bojar, 2011, Fishel et al., 2012, Gojun and Fraser, 2012]. For example, it is not unusual that an SMT system produces a translation with exactly the opposite meaning of the input (e.g. in double negation and negative concord phenomena).[7] Even from the statistical and engineering point of view, shallow SMT is inadequate because of its susceptibility to data sparseness problems and combinatorial explosion.

PB-SMT systems are rather good at lexical disambiguation in common phrases, partly thanks to the use of (surface string) n-gram language models. They, however, often fail to make the right lexical choice when the phrase is split into several parts or contains an uncommon word in the middle, which results in the necessary context being out of reach for an n-gram language model [Crego and Yvon, 2010]. On the other hand, SMT systems usually do not perform well in grammaticality of the target sentence. For highly inflecting languages, they are generally unable to model grammatical rules of the language, as compared to rule-based systems that typically employ a deterministic morphological generator and thus make significantly fewer errors in inflection [CeausuI and Tufis, 2011, Dove et al., 2012].

This problem can be partly diminished by employing a large language model, although long-distance dependencies and phrases containing uncommon words still continue to be inflected randomly rather than correctly [Guthrie et al., 2006, Wu and Matsumoto, 2014]. Also, as no fine-grained morphological tagging is usually employed, homonymous forms are not disambiguated and their agreement still cannot be easily captured [Minkov et al., 2007, Avramidis and Koehn, 2008, Rosa et al., 2012b].

Moreover, a further drawback of a powerful language model, especially when the translation model is much smaller (which is often the case), is the risk of misleading: a fluent output might not be a good translation of the source text. A strong language model can also easily force the translation model to leave out a word or to generate a surplus word, changing or even inverting the meaning of the sentence, or simply making the sentence completely unintelligible [Lambert et al., 2006]. A notable error of this type is leaving out the main predicate, which can easily happen as most SMT systems do not a have a way to tell that some words are more important than others [Ma and McKeown, 2009].

Both lexical and grammatical accuracy of the output is further hindered by the fact

[6] By *shallow SMT* we mean SMT which uses no linguistic analysis, or only shallow analysis: tokenization or morpheme splitting, tagging, lemmatization, but no syntax. Most current PB-SMT are shallow, though there are ways to use syntax (dependency or constituency) in PB-SMT, as discussed in Section 4.

[7] Errors in negation can mislead the users in a dangerous way because the translation may be fully grammatical and fluent, so the users may not notice this error (unlike with incomprehensible translations). For example, the Dutch phrase “TWV niet vereist” (work permit not required) is translated by Google Translate (as of May 2015) as “Permit required”. See `https://translate.google.com/#nl/en/TWV%20niet%20vereist` (note the missing period after the input phrase). Errors in negation are not uncommon in SMT [Bojar et al., 2013b, Fancellu and Webber, 2014], especially if one of the langauges uses double negation [Rosa, 2013, p. 25].

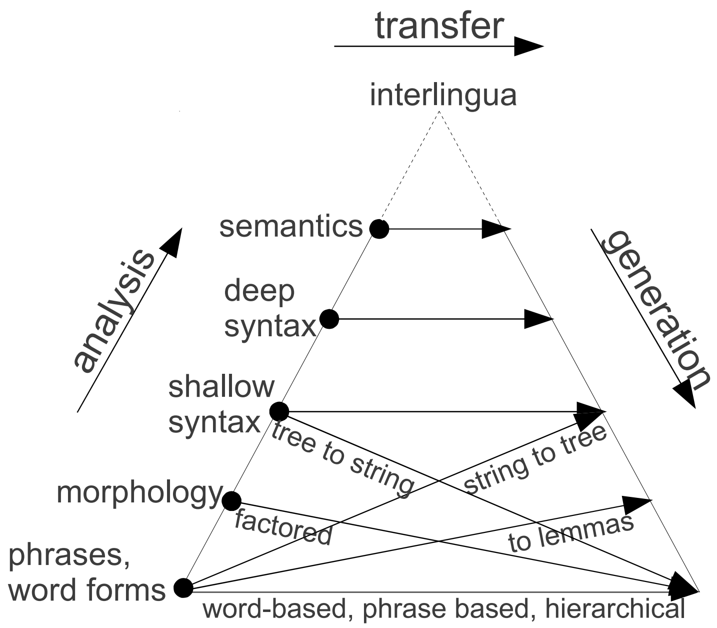

Figure 1: The Vauquois triangle.

that most SMT systems are unable to efficiently incorporate information from neighboring sentences or from the whole text and its domain, leading to errors in inter-sentential cohesion, broken coreference chains, domain-ignorant lexical choices, etc. [Hardmeier and Federico, 2010]. Punctuation is also often processed incorrectly, especially if the target language has clear but complex punctuation rules that the statistical models are unable to capture, or if the punctuation rules of the source and target language are substantially different [Koehn and Haddow, 2009].

4 Shallow Syntax-Based Approaches

In the previous section, we described approaches that operate on the lowest levels of the Vauquois triangle (see Figure 1), with minimal or no abstraction over the actual word forms. In this section, we will move upwards in the triangle.

Let us start with a terminological note: by syntax we mean linguistically-motivated syntax. Therefore, we do not consider hierarchical models described in Section 3.3 syntax-based, unless linguistically motivated non-terminals are used, e.g. as observed in a manually annotated treebank.

For **syntax-based SMT approaches**, several research directions have been proposed that use linguistically motivated structures on the source side only (tree-to-string), the target side only (string-to-tree), or both the source side and the target side (tree-to-tree).

4.1 Tree-to-string Translation

There are many ways of incorporating linguistically motivated syntax trees into machine translation.

Huang et al. [2006] is a prominent example of a tree-to-string translation approach called syntax-directed translation, in which the syntactic structure of the source sentence is used to guide the decoder in generating the target sentence. This work was further extended by Liu and Gildea [2008], who use not only the syntactic structure but also semantic role labels.

Zhang et al. [2007] combine the tree-to-string approach with hierarchical approach, using both linguistic and non-linguistic features.

4.2 String-to-tree Translation

The inverse approach, where the source sentence is used to generate the syntactic tree or tree fragments of the target sentence, is used less frequently. This approach is usually motivated by the effort to find a grammatically correct translation, especially when translating into a morphologically rich language. String-to-tree translation was pioneered by Galley et al. [2004] and later extended by Marcu et al. [2006] and Liu et al. [2007] by adjusting the parse trees for easier rule extraction. However, Williams and Koehn [2011] suggest that adding only a set of agreement constraints seems to be a more effective way of ensuring output grammaticality.

4.3 Tree-to-tree Translation

One of the most influential tree-to-tree approaches is **Synchronous tree substitution grammar** (STSG).[8] STSG was introduced into MT by Hajič et al. [2002], and formalized by Eisner [2003]. It is based on an assumption that a valid translation of an input sentence can be obtained by local structural changes of the input syntactic tree and translation of node labels while there exists a derivation process common to both of the languages. Although such an idea is theoretically promising and could remedy the above mentioned problems of PB-SMT, it has not outperformed PB-SMT in practice yet.

Bojar [2008] identified several causes of such low performance, one of which is an inherent STSG property that holds when applied to dependency trees: All immediate dependents of internal nodes must be included in a translation rule. This leads to data sparseness and prevents the extraction of many useful rules (e.g. seeing “he runs quickly” → “er läuft schnell” in the training data does not allow us to translate “he runs” unless we have also seen “he runs” → “er läuft” in the training data). Another problem of the STSG approach underlined by Bojar [2008] is that when a node with all its attributes (lemma and grammatical categories etc.) is taken as an atomic unit, the resulting translation model is very sparse and unreliable. When each attribute is translated separately using factored models but without taking other attributes into account, the resulting MT quality is unsatisfactory as well.

5 Post-editing

5.1 Human Post-editing

The highest quality of translations is still achieved by human translators. However, it has been shown that even when the translation is performed by a trained professional,

[8] STSG is a version of synchronous tree-adjoining grammar (STAG) [Shieber and Schabes, 1990] that lacks the adjunction operation. Unlike synchronous context-free grammar (SCFG) that are used e.g. in Hiero system mention in Section 3.3, STSG allows to model non-isomorphic pairs of trees because already the elementary trees (forming the STSG) may be non-isomorphic.

machine translation can be of benefit – Green et al. [2013] and other studies show that post-editing machine translation outputs by human translators leads to better and faster translations compared to unassisted translating. Such observations led to establishing the field of computer-aided translation, which deals with various ways in which MT systems and human translators may interact to achieve high-quality translations in an efficient way [Langlais et al., 2000, Koehn, 2009].

5.2 Automatic Post-editing

Observations from human post-editing led to the idea of automatic post-editing, which follows the usual practice of treating a difficult problem by breaking it down into several simpler subproblems. Moreover, post-editing is a common technique even with human translators and helps them in achieving high-quality translations. Therefore, it may be reasonable to assume that automatic post-editing is closer than one-step translation to the way that the highest quality translations are usually obtained.

The task of automatic post-editing (APE) is different from the typical task of one-step machine translation in some ways. Most of them stem from the fact that the post-editing system has access to a preliminary translation, which can be used both as a source of information for the system, and as a backoff output when the post-editing system is unable to improve upon the machine translation system. In the typical machine translation setting, effort is typically made to produce a translation even when the system is very unsure about its quality, so that out-of-vocabulary events are avoided. This is different with APE since the system should only make an edit if it is reasonably sure that it will improve the quality of the translation. This also allows for a simpler design of the APE system, as it need not have the ability to generate a full sentence – it suffices to spot and correct errors, leaving the rest of the sentence intact.

First APE approaches dealt with statistical post-editing of rule-based MT [Simard et al., 2007], and showed that the translation quality can be improved significantly using such approach. However, as SMT is the current state-of-the-art approach to machine translation, focus has moved to post-editing of SMT. Mildly positive results were reported for statistical post-editing of SMT by training an SMT system to post-edit its own outputs [Oflazer and Durgar El-Kahlout, 2007, Bechara et al., 2011].[9]

A more promising research path seems to be the use of an APE system that is substantially different from the first-stage SMT system, being strong in some aspects in which the SMT system is weak. First such approach is that of Stymne and Ahrenberg [2010], who use a grammar checker to correct errors in SMT from English to Swedish. Another such approach is the Depfix system [Rosa et al., 2012b] for English-to-Czech translation, which is described in the following section.

5.3 The Depfix System

Depfix performs rule-based post-editing of phrase-based SMT. The phrase-based system provides a baseline translation. Any phrase-based system can be used, even a closed-source one. The Depfix system analyzes both the source sentence and its translation provided by the PB-SMT system. In doing so, it employs various NLP tools, including morphological taggers and dependency parsers, which produce a feature-rich linguistic representation of

[9] Both Oflazer and Durgar El-Kahlout [2007] and Bechara et al. [2011] used only a small amount of training data, and it is probable that the method does not scale well.

the sentences. This representation is then used by a set of linguistic rules that detect many common mistakes made by PB-SMT systems (mostly grammatical mistakes, such as agreement errors). The detected errors are fixed in the linguistic representation, and a corrected translation is then produced using a morphological generator that generates the surface word forms from their linguistic representation. The Depfix system has been shown to be able to significantly improve the quality of outputs of most PB-SMT systems, including state-of-the-art systems.

Although probably constituting a promising research path, the current version of Depfix has many issues. The gravest problem is that the post-editing is implemented only for English-to-Czech translation, and, while being theoretically adaptable to other translation directions, the fact that Depfix is rule-based and the rules are manually crafted, makes any such adaptation laborious. Another drawback of Depfix is its inability to correct lexical errors, which often appear in MT systems' outputs and significantly lower their quality. It should also be noted that while benefiting from the level of abstraction provided by syntactic trees, Depfix is still unable to correctly capture many complex phenomena as a higher abstraction level would be needed for such task.

6 Deep-syntax-based Statistical Approaches

Some researchers are currently turning their attention to more complex SMT approaches that go beyond the basic syntax-based MT systems. Although it was thought to be an easy task at the beginning, machine translation has turned out to be a rather difficult one through the many years of research, and it seems reasonable to believe that we must leverage nearly all of our linguistic knowledge, resources, and tools to reach a satisfying quality of MT. This may combine phrase-based SMT approaches, which have proven to be very efficient and successful for MT, syntax-based SMT, which addresses issues that are extremely hard or impossible to handle by phrase-based SMT, deep language processing, which is able to abstract from semantically irrelevant surface differences, semantic analysis, which can capture the text meaning itself, but also rule-based approaches, which are extremely efficient at handling e.g. highly regular language features, etc.

In QTLeap, we specifically focus on *dependency* deep syntactic trees, as dependency trees are becoming more popular than constituency trees in many areas of computational linguistics in recent years. One of the important advantages of dependencies over constituencies is the fact that they can naturally capture some language phenomena common in the QTLeap languages – such as non-projective relations, which are frequent in languages with a relatively free word order. Dependencies are also the formalism used by most of the partners in all of their treebanks, deepbanks, and NLP tools, and the QTLeap members have therefore a strong expertise in dependency syntax.

6.1 Motivation for MT based on deep syntax

Efforts to build translation models around deep syntactic structure often move the level of linguistic abstraction a step deeper into semantic roles and relations, promising:

- Simplification of the transfer step because of a greater structural similarity between the deep structures of the two languages as compared to the surface phrases, even if we do not assume a language-independent representation.

- Better generalization. While phrase-based approaches operate on word forms and have to learn translations of all morphological variants of a word independently, transfer at a deep layer typically operates on lemmas (base forms of words) of content words. One may thus be able to achieve the same coverage using less parallel data, if morphological and surface-syntactic analysis and generation steps are handled by external tools or trained on (much larger) monolingual data. A further generalization happens for grammatical constructions, whose surface realizations are typically abstracted from, and only their meaning is captured by language-neutral attributes of the deep tree nodes: for instance, the passive and active verbal constructions can be represented nearly identically, differing only in the value of a single attribute.

- Improved grammaticality of the output. An indisputable advantage of the explicit representation of target-side sentence structure in MT is improved grammaticality of the MT output.

6.2 Deep MT prototype: TectoMT

While promising results have been reported on small training datasets [Hearne, 2005], most attempts to use deeper analysis in MT either have failed to match the phrase-based benchmark [Čmejrek et al., 2003, Žabokrtský et al., 2008, Bojar, 2008] or were never directly compared to the state-of-the-art [Richardson et al., 2001, Apresjan et al., 2003, Oepen et al., 2007]. One reason may be that MT's use of deeper analysis has had a different focus, aiming at higher output quality over system coverage [Riezler and Maxwell, 2006].

A notable exception is the TectoMT system [Žabokrtský et al., 2008, Žabokrtský and Popel, 2009] for English-to-Czech translation via deep syntactic representation. TectoMT regularly takes part in the WMT shared task on news text translation. In the last few years, even the manual evaluation has had troubles identifying whether TectoMT performs better or worse than the state-of-the-art phrase-based system trained on the same amount of data [Bojar et al., 2011].

The EuroMatrix project [Bojar and Týnovský, 2009] identified two main problems that have stood in the way of the potential deep analysis benefits — first, reduced training data due to the additional constraint of parallel syntactic structures, a problem also noticed in shallow syntactic approaches described in Section 4 [Chiang, 2010], and second, additional linguistic attributes due to each node in the deep representation having a few dozen detailed flags representing e.g. verbal tense, mood, or gender. Most of these attributes have to be provided by the transfer step to the final generator, which in the implementation of Bojar [2008] led to an unbearable memory cost (for all possible combinations of values of the attributes) or a high risk of search errors (due to heavy pruning of these combinations).

TectoMT approach avoids these traps by strictly decomposing trees into individual nodes (thus avoiding the data loss due to structural divergence) and by carefully handcrafted rules to transfer most of the attributes. The core of the TectoMT system is now a statistical transfer converting an English deep syntactic tree into the Czech counterpart [Žabokrtský and Popel, 2009]. It is only the lemmas of the words and so-called formemes, a very compact representation of morpho-syntactic relationships of neighboring words (e.g. the case and the preposition for nouns) that enter the search. The tree transfer in TectoMT relies on the assumption that the tree structure is isomorphic in the two languages. This assumption is not possible for surface-syntactic trees, but it is viable for the deep trees used in TectoMT, the so-called tectogrammatical trees as used in Prague dependency treebanks [Hajič et al., 2006, 2012].

7 Beyond the State of the Art

The key point of the QTLeap project in order to surpass the state-of-the-art in MT is to combine probabilistic and linguistically precise approaches in novel ways as this is becoming possible by recent advanced multilingual parallel DeepBanks and advanced statistical modeling techniques. This section presents just a reduced list of promising paths, taken from the QTLeap Description of Work, which will be described in future QTLeap deliverables.

Statistical methods and machine learning integrated with lexical semantic and deep processing will be used to get reliable deep structures of sentences (e.g. predicate-argument structure in tectogrammatical dependency trees with rich semantic attributes, MRS logical forms, etc.) This paves the way to change the current MT technological paradigm in several crucial aspects:

- by using tree-to-tree statistical transfer while extensively relying on deep linguistic representations that abstract away from language specific aspects and permit a quantum leap in terms of lowering data sparseness problems;
- by going beyond sentence boundaries; and
- by employing advanced and computationally-intensive machine learning techniques, including unsupervised methods exploiting very large corpora.

Tree-to-tree approaches have been used in previous generation systems that are based mainly on manually created rules (cf. Section 2). They are not yet fully explored though in the statistical framework (cf. Section 4.3) as opposed to current state-of-the-art string-to-string, string-to-tree and tree-to-string approaches. While doing so, we will use advanced statistical modeling techniques such as Hidden Tree Markov Models [Žabokrtský and Popel, 2009], CRF (Conditional Random Fields) or context-sensitive Maximum Entropy transfer models [Och and Ney, 2002, Mareček et al., 2010].

For feature extraction and annotation, we will extensively use existing deep resources, such as multilingual parallel DeepBanks [e.g., Flickinger et al., 2012].

We plan to go beyond sentence boundaries while building models for co-reference, document-level context features etc. The existing deep-syntactic MT system TectoMT[10] will be adapted to support non-isomorphic tree transfer and to exploit document-level context features and Linked Open Data.

For tree-to-tree transfer models, we will develop target-language tree models that improve grammatical and semantic cohesion of the translated sentences. Unlike n-gram language models, these models condition on the local tree context rather than local string context. Techniques inspired by generalized parallel backoff and factored language models [Bilmes and Kirchhoff, 2003] will be used to improve the translation and language model.

For language pairs where the resulting deep MT system will not outperform phrase-based systems significantly or where previous results show other promising directions for hybrid combinations of deep and stochastic information, we will try to combine different systems (our deep MT with mainstream SMT) and exploit the fact that they have different

[10] http://ufal.mff.cuni.cz/tectomt

distributions of errors. We also plan to extend the existing and successful Depfix system [Rosa et al., 2012a] that improves SMT output using automatic rule-based corrections of grammatical mistakes, and exploit also deep features in addition to the source and target language dependency trees.

8 Conclusion

Through many decades of research, machine translation has arrived to a state where complex phrase-based machine translation systems are considered to be the state-of-the-art approach to MT. They employ many additional modules that help them avoid a range of issues that lowered the quality of their outputs in the past, and typically achieve the best results among all types of MT systems [Callison-Burch et al., 2011, 2012, Bojar et al., 2013a]. However, the pace of improvements has slowed down recently, as if this approach has been depleted, and a new, different approach must be developed to replace PB-SMT, allowing further improvements to MT quality.

It seems that such an approach should be syntax-based to easily capture linguistic phenomena that are hard to address by PB-SMT, such as long-distance grammatical cohesion. Moreover, it is believed by some researchers that the typical syntactic approaches are too shallow for the task of MT, and therefore, methods of deep linguistic analysis should be used to capture all the important parts of the meaning of the text being translated.

References

Juan A Alonso and Gregor Thurmair. The Comprendium Translator system. In *Proceedings of the Ninth Machine Translation Summit*. International Association for Machine Translation (IAMT), 2003.

Jurij D Apresjan, Igor M Boguslavsky, Leonid L Iomdin, and Leonid L Tsinman. Lexical functions as a tool of ETAP-3. In *Proceedings of MTT 2003*, 2003.

Eleftherios Avramidis and Philipp Koehn. Enriching morphologically poor languages for statistical machine translation. In *ACL*, pages 763–770, 2008.

Hanna Bechara, Yanjun Ma, and Josef van Genabith. Statistical post-editing for a statistical MT system. In *Proceedings of the 13th Machine Translation Summit (MT Summit XIII)*, pages 308–315. International Association for Machine Translation, 2011. URL http://www.mt-archive.info/MTS-2011-Bechara.pdf.

Jeff A Bilmes and Katrin Kirchhoff. Factored language models and generalized parallel backoff. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003–short papers-Volume 2*, pages 4–6. Association for Computational Linguistics, 2003.

Igor Boguslavsky. A bi-directional Russian-to-English machine translation system (ETAP-3). *Proceedings of the Machine Translation Summit V. Luxembourg*, 1995.

Ondřej Bojar. *Exploiting Linguistic Data in Machine Translation*. PhD thesis, ÚFAL, MFF UK, Prague, Czech Republic, 2008.

Ondřej Bojar. Analyzing error types in english-czech machine translation. *The Prague Bulletin of Mathematical Linguistics*, 95:63–76, 2011.

Ondrej Bojar and Kamil Kos. 2010 failures in English-Czech phrase-based MT. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 60–66, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W10-1705.

Ondřej Bojar and Miroslav Týnovský. Evaluation of tree transfer system. Technical Report 3.4, Charles University in Prague, ÚFAL, 2009.

Ondřej Bojar, Miloš Ercegovčević, Martin Popel, and Omar Zaidan. A grain of salt for the WMT manual evaluation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 1–11, Edinburgh, Scotland, July 2011. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W11-2101.

Ondřej Bojar, Christian Buck, Chris Callison-Burch, Christian Federmann, Barry Haddow, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia. Findings of the 2013 Workshop on Statistical Machine Translation. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 1–44, Sofia, Bulgaria, August 2013a. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W13-2201.

Ondřej Bojar, Rudolf Rosa, and Aleš Tamchyna. Chimera – three heads for English-to-Czech translation. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 92–98, Sofia, Bulgaria, August 2013b. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W13-2208.

Francis Bond, Stephan Oepen, Eric Nichols, and Dan Flickinger. Deep open-source machine translation. *Machine Translation*, 25(2):87–105, June 2011.

Peter F. Brown, John Cocke, Stephen A. Della-Pietra, Vincent J. Della-Pietra, Frederick Jelinek, Robert L. Mercer, and Paul Rossin. A statistical approach to language translation. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, 1988.

Peter F. Brown, John Cocke, Stephen A. Della-Pietra, Vincent J. Della-Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul Rossin. A statistical approach to machine translation. *Computational Linguistics*, 16(2):76–85, 1990.

Peter F. Brown, Stephen A. Della-Pietra, Vincent J. Della-Pietra, and Robert L. Mercer. The mathematics of statistical machine translation. *Computational Linguistics*, 19(2): 263–313, 1993.

Chris Callison-Burch, Philipp Koehn, Christof Monz, and Omar Zaidan. Findings of the 2011 workshop on statistical machine translation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 22–64, Edinburgh, Scotland, July 2011. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W11-2103.

Chris Callison-Burch, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia. Findings of the 2012 workshop on statistical machine translation. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 10–48, Montreal, Canada, June 2012. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W12-3102.

Alexandru CeausuI and Dan Tufis. Addressing smt data sparseness when translating into morphologically-rich languages. In *Human-Machine Interaction in Translation: Proceedings of the 8th International NLPCS Workshop*, volume 41, page 57. Samfundslitteratur, 2011.

Mauro Cettolo, Marcello Federico, Daniele Pighin, and Nicola Bertoldi. Shallow-syntax phrase-based translation: Joint versus factored string-to-chunk models. In *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas (AMTA)*, pages 56–64, 2008.

David Chiang. Learning to translate with source and target syntax. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1443–1452, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P10-1146.

David Chiang, Adam Lopez, Nitin Madnani, Christof Monz, Philip Resnik, and Michael Subotin. The Hiero machine translation system: Extensions, evaluation, and analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 779–786, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/H05-1098.

Josep Maria Crego and François Yvon. Improving reordering with linguistically informed bilingual n-grams. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 197–205, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL http://dl.acm.org/citation.cfm?id=1944566.1944589.

Catherine Dove, Olga Loskutova, and Ruben de la Fuente. What's your pick: Rbmt, smt or hybrid? In *Proceedings of the tenth conference of the Association for Machine Translation in the Americas (AMTA 2012). San Diego, CA*, 2012.

Jordi Duran, Lluís Villarejo, Mireia Farrús, Sergio Ortiz, and Gema Ramírez. Incorporating subject areas into the apertium machine translation system. In Adam Przepiórkowski, Maciej Piasecki, Krzysztof Jassem, and Piotr Fuglewicz, editors, *Computational Linguistics*, volume 458 of *Studies in Computational Intelligence*, pages 281–292. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-34398-8. doi: 10.1007/978-3-642-34399-5_15. URL http://dx.doi.org/10.1007/978-3-642-34399-5_15.

Chris Dyer, Adam Lopez, Juri Ganitkevitch, Jonathan Weese, Ferhan Ture, Phil Blunsom, Hendra Setiawan, Vladimir Eidelman, and Philip Resnik. cdec: A decoder, alignment, and learning framework for finite-state and context-free translation models. In *Proceedings of the ACL 2010 System Demonstrations*, pages 7–12, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P10-4002.

Jason Eisner. Learning non-isomorphic tree mappings for machine translation. In Yuji Matsumoto, editor, *The Companion Volume to the Proceedings of 41st Annual Meeting of the Association for Computational Linguistics*, pages 205–208, 2003. URL http://www.aclweb.org/anthology/P03-2041.pdf.

Federico Fancellu and Bonnie Webber. Applying the semantics of negation to SMT through n-best list re-ranking. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 598–606, Gothenburg, Sweden, April 2014. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/E14-1063.

Christian Federmann and Sabine Hunsicker. Improving an existing RBMT system by stochastic analysis. In *Hamburg Working Paper in Multilingualism*, Hamburg Working Paper in Multilingualism. German Society for Computational Linguistics and Language Technology, 2011a.

Christian Federmann and Sabine Hunsicker. Stochastic parse tree selection for an existing RBMT system. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 351–357. Association for Computational Linguistics, 2011b.

Mark Fishel, Ondrej Bojar, and Maja Popovic. Terra: a collection of translation error-annotated corpora. In *LREC*, pages 7–14, 2012.

Daniel Flickinger, Yi Zhang, and Valia Kordoni. DeepBank: A dynamically annotated treebank of the Wall Street Journal. In *Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories*, pages 85–96, 2012.

Ariadna Font Llitjós and Stephan Vogel. A walk on the other side: Using SMT components in a transfer-based translation system. In *Proceedings of SSST, NAACL-HLT 2007 / AMTA Workshop on Syntax and Structure in Statistical Translation*, pages 72–79, Rochester, New York, April 2007. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W07-0410.

Mikel L. Forcada, Mireia Ginestí-Rosell, Jacob Nordfalk, Jim O'Regan, Sergio Ortiz-Rojas, JuanAntonio Pérez-Ortiz, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez, and Francis M. Tyers. Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation*, 25(2):127–144, 2011. ISSN 0922-6567. doi: 10.1007/s10590-011-9090-0. URL http://dx.doi.org/10.1007/s10590-011-9090-0.

Alexander Fraser, Marion Weller, Aoife Cahill, and Fabienne Cap. Modeling inflection and word-formation in SMT. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 664–674, Avignon, France, April 2012. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/E12-1068.

Michel Galley and Christopher D. Manning. Accurate non-hierarchical phrase-based translation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 966–974, Los Angeles, California, June 2010. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/N10-1140.

Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. What's in a translation rule? In *Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL)*, 2004.

Anita Gojun and Alexander Fraser. Determining the placement of german verbs in english-to-german smt. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 726–735, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics. ISBN 978-1-937284-19-0. URL http://dl.acm.org/citation.cfm?id=2380816.2380905.

Spence Green, Jeffrey Heer, and Christopher D. Manning. The efficacy of human post-editing for language translation. In *ACM Human Factors in Computing Systems (CHI)*, 2013. URL http://vis.stanford.edu/papers/post-editing.

David Guthrie, Ben Allison, Wei Liu, Louise Guthrie, and Yorick Wilks. A closer look at skip-gram modelling. In *Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006)*, pages 1–4, 2006.

Nizar Habash and Bonnie J. Dorr. Handling translation divergences: Combining statistical and symbolic techniques in generation-heavy machine translation. In Stephen D. Richardson, editor, *Machine Translation: From Research to Real Users, 5th Conference of the Association for Machine Translation in the Americas, AMTA 2002 Tiburon, CA,* *USA, October 6-12, 2002, Proceedings*, volume 2499 of *Lecture Notes in Computer Science*. Springer, 2002. ISBN 3-540-44282-0.

Jan Hajič, Martin Čmejrek, Jason Eisner, Gerald Penn, Owen Rambow, Drago Radev, Yuan Ding, Terry Koo, and Kristen Parton. Natural language generation in the context of machine translation. Technical Report 1, CLSP JHU, USA, 2002.

Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, and Magda Ševčíková-Razímová. Prague dependency treebank 2.0, 2006.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. Announcing Prague Czech-English dependency treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 3153–3160, İstanbul, Turkey, 2012. ELRA, European Language Resources Association. ISBN 978-2-9517408-7-7.

Christian Hardmeier and Marcello Federico. Modelling pronominal anaphora in statistical machine translation. In *Proceedings of the 7th International Workshop on Spoken Language Translation, Paris. :*, pages 283–289, 2010.

Mary Hearne. *Data-oriented models of parsing and translation.* PhD thesis, Dublin City University, 2005.

Mark Hopkins and Jonathan May. Tuning as ranking. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1352–1362, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/D11-1125.

Liang Huang, Kevin Knight, and Aravind Joshi. A syntax-directed translator with extended domain of locality. In *Proceedings of the Workshop on Computationally Hard Problems and Joint Inference in Speech and Language Processing*, pages 1–8, New York City, New York, June 2006. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W06-3601.

Philipp Koehn. A process study of computer-aided translation. *Machine Translation*, 23 (4):241–263, November 2009.

Philipp Koehn and Barry Haddow. Edinburgh's submission to all tracks of the wmt2009 shared task with reordering and speed improvements to moses. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, StatMT '09, pages 160–164, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. URL http://dl.acm.org/citation.cfm?id=1626431.1626464.

Philipp Koehn and Hieu Hoang. Factored translation models. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 868–876, 2007. URL http://www.aclweb.org/anthology/D07-1091.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase based translation. In *Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL)*, 2003. URL http://acl.ldc.upenn.edu/N/N03/N03-1017.pdf.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Christopher J. Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P07-2045.

Patrik Lambert, Jesús Giménez, Marta R Costa-Jussà, Enrique Amigó, Rafael E Banchs, Lluís Márquez, and JAR Fonollosa. Machine translation system development based on human likeness. In *Spoken Language Technology Workshop, 2006. IEEE*, pages 246–249. IEEE, 2006.

Philippe Langlais, George Foster, and Guy Lapalme. Transtype: a computer-aided translation typing system. In *Proceedings of the ANLP-NAACL 2000 Workshop on Embedded Machine Translation Systems*, 2000.

Zhifei Li, Chris Callison-Burch, Chris Dyer, Sanjeev Khudanpur, Lane Schwartz, Wren Thornton, Jonathan Weese, and Omar Zaidan. Joshua: An open source toolkit for parsing-based machine translation. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 135–139, Athens, Greece, March 2009. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W09-0424.

Ding Liu and Daniel Gildea. Improved tree-to-string transducer for machine translation. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 62–69, Columbus, Ohio, June 2008. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W08-0308.

Yang Liu, Yun Huang, Qun Liu, and Shouxun Lin. Forest-to-string statistical translation rules. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 704–711, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P07-1089.

Wei-Yun Ma and Kathleen McKeown. Where's the verb?: Correcting machine translation during question answering. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 333–336, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. URL http://dl.acm.org/citation.cfm?id=1667583.1667686.

Daniel Marcu, Wei Wang, Abdessamad Echihabi, and Kevin Knight. SPMT: Statistical machine translation with syntactified target language phrases. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 44–52, Sydney, Australia, July 2006. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W06-1606.

David Mareček, Martin Popel, and Zdeněk Žabokrtský. Maximum entropy translation model in dependency-based mt framework. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 201–206, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W10-1730.

Einat Minkov, Kristina Toutanova, and Hisami Suzuki. Generating complex morphology for machine translation. In *ACL*, volume 7, pages 128–135, 2007.

Franz Josef Och. Minimum error rate training in statistical machine translation. In Erhard Hinrichs and Dan Roth, editors, *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 160–167, 2003. URL http://www.aclweb.org/anthology/P03-1021.pdf.

Franz Josef Och and Hermann Ney. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting of the Association of Computational Linguistics (ACL)*, 2000. URL http://acl.ldc.upenn.edu/P/P00/P00-1056.pdf.

Franz Josef Och and Hermann Ney. Discriminative training and maximum entropy models for statistical machine translation. In *Proceedings of the 40th Annual Meeting of the Association of Computational Linguistics (ACL)*, 2002.

Franz Josef Och and Hans Weber. Improving statistical natural language translation with categories and rules. In *Proceedings of the 36th Annual Meeting of the Association of Computational Linguistics (ACL)*, 1998.

Franz Josef Och, Christoph Tillmann, and Hermann Ney. Improved alignment models for statistical machine translation. In *Proceedings of the Joint Conference of Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP-VLC)*, pages 20–28, 1999.

Stephan Oepen, Erik Velldal, Jan Tore Lønning, Paul Meurer, Victoria Rosén, and Dan Flickinger. Towards hybrid quality-oriented machine translation. on linguistics and probabilities in MT. In *11th International Conference on Theoretical and Methodological Issues in Machine Translation*, 2007.

Kemal Oflazer and Ilknur Durgar El-Kahlout. Exploring different representational units in English-to-Turkish statistical machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 25–32, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W07-0204.

Stephen D Richardson, William B Dolan, Arul Menezes, and Monica Corston-Oliver. Overcoming the customization bottleneck using example-based MT. In *Proceedings of the workshop on Data-driven methods in machine translation-Volume 14*, pages 1–8. Association for Computational Linguistics, 2001.

Stefan Riezler and John T. Maxwell. Grammatical machine translation. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pages 248–255, New York City, USA, June 2006. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/N06-1032.

Rudolf Rosa. Automatic post-editing of phrase-based machine translation outputs. Master's thesis, ÚFAL, MFF UK, Prague, Czech Republic, 2013. URL http://ufal.mff.cuni.cz/%7Erosa/dp/rosa_dp_depfix.pdf.

Rudolf Rosa, Ondřej Dušek, David Mareček, and Martin Popel. Using parallel features in parsing of machine-translated sentences for correction of grammatical errors. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 39–48, Jeju, Republic of Korea, July 2012a. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W12-4205.

Rudolf Rosa, David Mareček, and Ondřej Dušek. Depfix: A system for automatic correction of Czech MT outputs. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 408–414, Montreal, Canada, June 2012b. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W12-3151.

Vícor M. Sánchez-Cartagena, Miquel Esplá-Gomis, Felipe Sánchez-Martínez, and Juan Antonio Pérez-Ortiz. Choosing the correct paradigm for unknown words in rule-based machine translation systems. In *Proceedings of the Third International Workshop on Free/Open-Source Rule-Based Machine Translation*, pages 27–39, Gothenburg, Sweden, June 2012.

Stuart M. Shieber and Yves Schabes. Synchronous tree-adjoining grammars. In *COLING*, pages 253–258, 1990.

Michel Simard, Cyril Goutte, and Pierre Isabelle. Statistical phrase-based post-editing. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 508–515, Rochester, New York, April 2007. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/N07-1064.

Sara Stymne and Lars Ahrenberg. Using a grammar checker for evaluation and postprocessing of statistical machine translation. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010. European Language Resources Association (ELRA). ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/pdf/426_Paper.pdf.

Peter Toma. Systran as a multilingual machine translation system. In *Proceedings of the Third European Congress on Information Systems and Networks, Overcoming the language barrier*, pages 569–581, 1977.

Kristina Toutanova, Hisami Suzuki, and Achim Ruopp. Applying morphology generation models to machine translation. In *Proceedings of ACL-08: HLT*, pages 514–522, Columbus, Ohio, June 2008. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P08-1059.

Francis Tyers, Felipe Sánchez-Martínez, Sergio Ortiz-Rojas, and Mikel Forcada. Free/open-source resources in the apertium platform for machine translation research and development. *The Prague Bulletin of Mathematical Linguistics*, 93:67–76, 2010.

Francis M. Tyers, Felipe Sánchez-Martínez, and Mikel L. Forcada. Unsupervised training of maximum-entropy models for lexical selection in rule-based machine translation. In *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*, pages 145–152, Antalya, Turkey, May 2015.

Martin Čmejrek, Jan Cuřín, and Jiří Havelka. Czech-English dependency tree-based machine translation. In *Proceedings of Meeting of the European Chapter of the Association of Computational Linguistics (EACL)*, 2003.

David Vilar, Daniel Stein, Matthias Huck, and Hermann Ney. Jane: Open source hierarchical translation, extended with reordering and lexicon models. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 268–276, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W10-1739.

Stephan Vogel, Hermann Ney, and Christoph Tillmann. HMM-based word alignment in statistical translation. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, 1996.

Rui Wang, Petya Osenova, and Kiril Simov. Linguistically-augmented bulgarian-to-english statistical machine translation model. In *Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra)*, pages 119–128, Avignon, France, April 2012a. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W12-0116.

Rui Wang, Petya Osenova, and Kiril Simov. Linguistically-enriched models for Bulgarian-to-English machine translation. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 10–19, Jeju, Republic of Korea, July 2012b. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W12-4202.

Taro Watanabe, Hajime Tsukada, and Hideki Isozaki. NTT system description for the WMT2006 shared task. In *Proceedings on the Workshop on Statistical Machine Translation*, pages 122–125, New York City, June 2006. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W06-3115.

Philip Williams and Philipp Koehn. Agreement constraints for statistical machine translation into German. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 217–226, Edinburgh, Scotland, July 2011. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W11-2126.

Xiaoyi Wu and Yuji Matsumoto. A hierarchical word sequence language model. In *Proceedings of the 28th Pacific Asia Conference on Language, Information, and Computation*, pages 489–494, Phuket,Thailand, December 2014. Department of Linguistics, Chulalongkorn University. URL http://www.aclweb.org/anthology/Y14-1056.

Zdeněk Žabokrtský and Martin Popel. Hidden Markov tree model in dependency-based machine translation. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 145–148, Suntec, Singapore, August 2009. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P09-2037.

Zdeněk Žabokrtský, Jan Ptáček, and Petr Pajas. TectoMT: Highly modular MT system with tectogrammatics used as transfer layer. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 167–170, Columbus, Ohio, June 2008. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W08-0325.

Dongdong Zhang, Mu Li, Chi-Ho Li, and Ming Zhou. Phrase reordering model integrating syntactic knowledge for SMT. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 533–540, 2007. URL http://www.aclweb.org/anthology/D07-1056.