

# Master's Thesis Review

Author: Amir Kamran B. Sc.  
Title: Hybrid Machine Translation Approaches for Low-Resource Languages  
Supervisor: Mgr. Martin Popel

The goal of Amir Kamran's thesis is to improve English-to-Urdu machine translation using source side preprocessing. The MT system used is Moses which also serves as the baseline. The author uses manual rules of Jawaid [2010] for changing SOV to SVO word order. As an alternative, he introduces an algorithm for automatic extraction of such rules given source-side parse trees (created by Stanford parser), target-side strings and a word alignment. For EMILLE English-Urdu corpus this approach proved to be significantly better than the baseline and comparable to the manual rules. For CRULP Penn Treebank, there was no significant improvement over the baseline. The second presented experiment is adding artificial post-markers to the source text, so it can be better aligned (by GIZA++) to Urdu. Although the post-markers subjectively seem to improve the alignment, no significant improvement in translation quality was measured.

Chapter 2 presents a background about machine translation, reordering techniques and evaluation metrics. Chapter 3 describes the corpora and tools used. Chapter 4 presents a very nice and illustrative description of post-markers in Urdu, factor-based models and it reports the experimental results with post-markers. The experiments with manual and automatic reordering rules are described in Chapter 5. Chapter 6 concludes.

The text is well structured and mostly easy to follow. Typographical quality is sufficient. There is a list of tables and figures and the attached CD contains all source codes as well as the data used. All experiments are appropriately documented and the results are discussed. I appreciate rigorous application of significance testing which revealed some of the achieved improvements to be not significant, which is correctly reflected in the discussions.

However, I have detected also the following flaws in the thesis:

- Whole three paragraphs of Section 2.7.4 (20 lines of text) are almost exact copy of Zhang, Vogel and Waibel [2004]. The usage of citation is indicated in the text only by the last sentence: “*Instead, we use the well-known bootstrapping technique to measure the confidence interval for BLEU/NIST [Koehn, 2004, Zhang and Vogel, 2004]*” (sic! – missing Waibel). In fact, also this last sentence is a citation of Zhang, Vogel and Waibel [2004] – except for the references, of course. Similarly, the following two paragraphs of Section 2.7.4 (15 lines of text) are almost exact copy of Koehn [2004], which is indicated just by the first sentence: “*Koehn [2004] describe (sic!) a non-analytical method to compute confidence intervals for BLEU/NIST metrics, called bootstrap resampling.*”  
On several places, slightly reworded sentences from Visweswariah et al. [2010] are used without references to it. For example, the original paper contains a sentence: “*Other models (Tillman, 2004), (Al-Onaizan and Papineni, 2006) generalize this to include lexical dependencies on the source.*” while Section 2.6 contains: “*Some later models generalize the distance based distortion model to include lexical independencies on the source (Tillmann, 2004; Koehn, et. al, 2005; Al-Onaizan and Papineni, 2006) by applying different weights to different phrases.*” The usage of *word independencies* does not make sense in this context.  
The whole first paragraph of Section 5.1 is rephrasing Visweswariah et al. [2010] (substituting *Hindi* for *Urdu*), but only the penultimate sentence contains a reference to it.
- The included CD contains source codes and a Makefile for reproducing the reported results. Unfortunately, the scripts contain absolute paths and several other errors, so it actually does not work outside author's own environment. Some source codes and scripts are taken from other tools (e.g. Moses, Stanford Parser), some are written by Amir Kamran, but they do not contain any information about the authorship. Neither it is mentioned in the thesis text which source codes are written by the author.
- Chapter 4 says: “*We will compare the suggested approach with factored based model for English-to-Urdu.*” Factored models are described in Section 4.2, but there is no comparison presented, and actually it seems that the author has not run the factored translation.

Also, Section 4.2 Factor-Based Models says: “*Jawaid [2010] showed improvements for English-to-Urdu.*” As far as I know, Jawaid [2010] presents no improvements with the factor-based approach. Otherwise, it would be logical to mention these results and use them as a secondary baseline in Section 4.4. Comparison with various settings of factors in factor-based translation would be beneficial.

- The manual reordering rules described in Section 5.1 are said to be “*a source-side reordering technique based on hand written tree transformation rules [Jawid, 2010].*” (sic! should be [Jawaid, 2010]). Also the two corpora used (EMILLE and Penn) are said to be “*modified and cleaned by Jawaid [2010] and we are using the modified versions.*” However, the BLEU scores reported in Table 5.2 are notably lower than in Jawaid [2010] – for EMILLE 20.47 vs. 25.15, for Penn 19.69 vs. 24.07. What is the reason?
- In Chapters 2 and 3, there is too much space (relatively to the length of the thesis) devoted to describing well-known facts and techniques, which are just loosely related to the main focus of the thesis. The description is sometimes repeating the same fact again and sometimes it is slightly inaccurate. I think it would be enough to write just few sentences about it, plus a standard reference to the full description of BLEU, NIST, bootstrapping confidence intervals, Moses etc. More importantly, there is almost no description of (and comparison with) the most relevant related work of Visweswariah et al. – just one sentence in Section 2.6 and one in Section 5.2.
- According to the author, “*manual reordering rules are developed with a lot of efforts and with cumbersome analysis of both the languages*” but the automatic rules led to a results comparable to manual rules for the EMILLE corpus (though significantly worse for the Urdu Penn Treebank). This result is promising, but its utility is questionable, because the presented approach is not fully automatic – it needed 100 manually aligned sentences, which is not a lot of human work. However, it is a question whether it would not be easier to manually create simple reordering rules similar to the 55 automatically learned ones (listed in Appendix A). Using automatic instead of manual alignment is one experiment that should be performed. Another idea is to select the 100 sentences in such a way that more linguistic phenomena are cover compared to the random selection. This can improve the results.
- The abstract is an exact copy of the assignment of the master thesis, but it should not be (see [http://www.mff.cuni.cz/toISO-8859-2/en/studium/bcmgr/prace/dp\\_uprava\\_en.pdf](http://www.mff.cuni.cz/toISO-8859-2/en/studium/bcmgr/prace/dp_uprava_en.pdf)). The abstract mentions “*dependency-based reordering*”, “*Automatic extraction of lexical and syntactic rules using statistical methods to facilitate the Transfer-Based Machine Translation*” and “*lexical and syntactic rules to build a rule-based machine translation system*”, but these topics are not covered in the thesis. On the other hand, the abstract does not mention markers described in Chapter 4.
- References are inconsistent in style, and more importantly some required information is missing – e.g. one paper is cited as “*Lav, 2004*” on page 1, but the author is missing in the References section (it is Alon Lavie et al.), so neither the name of the paper can be found by uninformed readers. Similarly for “*Bak, 2002*” on page 13.
- For each of the four experiments (baseline, markers, manual reordering, automatic reordering), there are two separate tables – one with BLEU & NIST scores (e.g. Table 5.5) and one with BLEU confidence intervals (e.g. Table 5.6) which could be merged. Also, from my point of view, it would be more interesting to see various methods compared in one table than various treebanks.
- Errors in English grammar (e.g. “*Chapter 2 provide*”), typos (e.g. *SBMT* instead of *EBMT*) and other errors (e.g. “*Section ?? discuss*”) are quite frequent (about two errors per page on average).

English-to-Urdu machine translation is a difficult task, because Urdu is a low-resource language. The experiment with automatic reordering rules showed an improvement over the baseline without using manual rules. The experiment with post-markers is well-motivated and it is a promising ground for future work. Therefore, I think the main goal of the thesis was achieved and I recommend this thesis for the defense.

Prague, August 2011

Mgr. Martin Popel  
ÚFAL MFF UK