

---

# Response-Based Learning for Patent Translation

**Stefan Riezler**

Computational Linguistics & IWR  
Heidelberg University, 69120 Heidelberg, Germany

riezler@cl.uni-heidelberg.de

---

## Abstract

In response-based structured prediction, instead of a gold-standard structure, the learner is given a response to a predicted structure from which a supervision signal for structured learning is extracted. Applied to statistical machine translation (SMT), different types of environments such as a downstream application, a professional translator, or an SMT user, may respond to predicted translations with a ranking, a correction, or an acceptance/rejection decision, respectively. We present algorithms and experiments that show that learning from responses alleviates the supervision problem and allows a direct optimization of SMT for tasks such as cross-lingual patent prior art retrieval, or translation of technical patent documents.

## 1 Introduction

*Response-based learning* describes a range of statistical learning methods that replace the full-information supervised learning scenario by extracting supervision signals from the response of an extrinsic environment to a predicted translation. Learning proceeds by “trying out” or “grounding” translations in a task that is external to translation itself, receiving a response from interacting in this task, and converting this response into a supervision signal for updating model parameters. We focus on response-based structured prediction that operates according to the following learning protocol:

1. Environment generates input structure  $x_t$
2. Learner *predicts* output structure  $y_t$
3. Environment generates *response* signal  $r_t$
4. Learner uses pair  $(x_t, r_t)$  to *update* its prediction rule

Clearly, the key advantage of response-based learning is to alleviate the supervision problem by a repeatable generate-and-test procedure where feedback is obtained from the environment. Such feedback is in general easier and less costly to obtain than full supervision information. Furthermore, learning from task-based responses has the effect of grounding the learning process in an extrinsic environment.

In this paper, we will describe three different types of response signals that are elicited by grounding SMT into three different environments, namely *ranking* responses elicited in *cross-lingual information retrieval* (Section 2), *correction* responses elicited in form of *post-edits* by *professional translators* (Section 3), and *acceptance/rejection* responses elicited in *personalized SMT* (Section 4).

## 2 Grounding SMT in Cross-Lingual Patent Retrieval

The industry standard in cross-lingual information retrieval (CLIR) is to use established translation models for context-aware translation of query strings, effectively reducing the problem

	Wikipedia			Patents		
	MAP	NDCG	PRES	MAP	NDCG	PRES
DT	.3678	.5691	.7219	.2554	.5397	.5680
PSQ	.3642	.5671	.7165	.2659	.5508	.5851
BOW-FD	*.3919	*.5963	*.7528	*.2883	*.5819	*.6251

Table 1: Retrieval results of baseline systems and BOW-FD on large-scale CLIR task. \* denotes significant difference compared to both baselines.

of CLIR to a pipeline of direct translation (DT) and monolingual retrieval (Chin et al., 2008). Only recently, research approaches to probabilistic structured queries (PSQ) have been presented, that is, they include (weighted) translation alternatives into the query structure to allow a more generalized term matching (Ture et al., 2012a,b). In both DT and PSQ, the integration of SMT remains agnostic about its use for CLIR, and is instead optimized to match fluent, human reference translations. In contrast, retrieval systems often use bag-of-word representations, stopword filtering, and stemming techniques during document scoring, and queries are rarely fluent, grammatical natural language queries (Downey et al., 2008).

Attempts to inform the SMT system about its use for retrieval by optimizing its parameters towards a retrieval objective have been presented in the form of re-ranking (Nikoulina et al., 2012) or ranking (Sokolov et al., 2014).

The most direct integration of SMT and CLIR has been presented by Hieber and Riezler (2015) in an approach to *Bag-of-Words Forced Decoding* (BOW-FD) where IR features for words in the bag-of-words representation of documents force the SMT decoder to prefer relevant documents with high probability. The crucial steps of the response-based online learning protocol are instantiated in this approach as follows:

**Prediction:** Full translation hypergraph.

**Response:** BM25 scores of partial translation hypotheses.

**Learning:** Jointly optimize SMT and IR feature weights by direct ranking optimization on relevant documents.

The key advantage of this approach clearly is the exploitation of the full translation search space, which is made possible by decomposable IR features that relate partial translation hypotheses to documents in the retrieval collection. This in turn allows to use the SMT decoder directly in retrieval. The key modeling idea is to factorize the cross-lingual search problem of finding document  $d_e$  given query  $q_f$  as follows:

$$P(d_e|q_f) = \sum_h \underbrace{P(h|q_f)}_{\text{translation}} \times \underbrace{P(d_e|h, q_f)}_{\text{retrieval}}$$

Learning is done by a joint optimization of a score function for linear models  $F_{smt}$  and  $F_{ir}$  where

$$score(q_f, d_e) = \max_h e^{F_{smt}(h, q_e, q_f) + F_{ir}(h, d_e)}.$$

The parameters of this objective are optimized for ranking for given relevance ranked documents s.t.

$$rank(d^+, q) > rank(d^-, q) \iff score(d^+, q) > score(d^-, q).$$

Table 1 presents evaluation results for large-scale patent CLIR. It shows a clear advantage for BOW-FD over the aforementioned DT and PSQ approaches. Data used were Japanese-

	Patents	
	BLEU	$\Delta[\sigma]$
baseline	30.26	
rerank	32.54	+2.28 [ $\pm 1.47$ ]
tm+lm	33.24	+2.98 [ $\pm 2.03$ ]
tm+lm+rerank	<b>34.02</b>	+3.76 [ $\pm 2.08$ ]

Table 2: BLEU scores over patent test documents and mean difference and standard deviation from baseline (in small font size).

English Patent CLIR data<sup>1</sup> and German-English Wikipedia CLIR data<sup>2</sup>. Both datasets were extracted automatically by using the citation graph in patents and Wikipedia to extract ranked relevance links.

For more information on the data, model, and learning of BOW-FD, see Hieber and Riezler (2015).

### 3 Learning SMT from Translator Post-Edits

Recent research in computer-assisted translation (CAT) has shown that post-editing of machine translations by professional translators leads to improved productivity of translators, and to improved quality of final translations (Koehn, 2009; Garcia, 2011; Green et al., 2013). This scenario can be turned on its head by focusing on the human post-editor supporting the SMT system, leading to a mutually beneficial cycle of *human-assisted machine translation* where the SMT system performs online learning from human post-edits. The goal is to improve translation consistency for a given document, and to offer the user the experience of a system that immediately learns from corrections (Wäschle et al., 2013; Denkowski et al., 2014; Green et al., 2014).

Patent data are especially well-suited for an application of this scenario because of the high repetitiveness of patent documents. Wäschle et al. (2013) and Bertoldi et al. (2014) recently presented an application of human-assisted SMT to patent data and data from legal and IT domains.

The crucial steps of the response-based online learning protocol are instantiated in online learning from post-edits as follows:

**Prediction:** Most probable sentence translation.

**Response:** User post-edit.

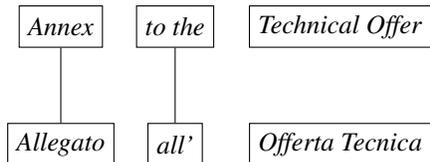
**Learning:** Dynamically extend phrase table and language model; update feature weights by reranking.

The key difference between online learning from post-edits instead of from reference translations is the dynamic extension of the phrase table. Wäschle et al. (2013) and Bertoldi et al. (2014) use a constrained search technique (Cettolo et al., 2010) that optimizes the coverage of both source and target sentences. It produces exactly one phrase segmentation and alignment, and allows gaps such that some source and target words may be uncovered. It differs in this respect from forced decoding which produces an alignment only when the target is fully reachable with the given models.

Assume the following phrase segmentation and alignment:

<sup>1</sup>111k train + 1,088k test, available under [www.cl.uni-heidelberg.de/boostclir](http://www.cl.uni-heidelberg.de/boostclir)

<sup>2</sup>245k train + 1,455k test, available under [www.cl.uni-heidelberg.de/wikiclr](http://www.cl.uni-heidelberg.de/wikiclr)



From this, three types of phrase pairs can be collected: (i) new phrase pairs by aligning unambiguous gaps (*Technical Offer* → *Offerta Tecnica*); (ii) known phrase pairs already present in the given model (*Annex* → *Allegato* and *to the* → *all*); (iii) full phrase pairs consisting of the complete source sentence and its user translation (*Annex to the Technical Offer* → *Allegato all' Offerta Tecnica*). Only phrases that contain at least one content word are considered.

The weight update for the extended phrase table is done by updating relative frequency features (collected into a cache during online learning) using a perceptron update on a training example  $(x_t, y_t)$  with feature representation  $f(x_t, y_t)$  in case the prediction  $\hat{y} = \arg \max_y w^\top f(x_t, y)$  does not match the target  $y_t$

$$w = w + f(x_t, y_t) - f(x_t, \hat{y}).$$

Table 2 shows experimental results for German-English patent data sampled from title, abstract and description sections from the PatTR corpus.<sup>3</sup> Significant gains in BLEU score can be obtained by reranking using a perceptron, or by updating phrase table (tm) and language model (lm). In combination, nearly 4 BLEU points improvement are achieved.

For more information on data, model, and learning, see Wäschle et al. (2013) and Bertoldi et al. (2014).

#### 4 Towards Personalized SMT: Learning from Partial User Feedback

While the above described post-editing scenario is a big step towards high-quality human-assisted SMT, the high cost and effort of post-editing may dampen the excitement about this scenario. The question to ask is whether production-quality post-edits from professional translators are really necessary for human-assisted SMT, or whether weaker feedback from less specialized users might be sufficient for learning.

Sokolov et al. (2014) recently addressed this question from a *coactive learning* perspective that provides a formal notion of feedback strength. They present a convergence analysis of online structured prediction algorithms that learn from feedback consisting of slight improvements over predicted translations, instead of optimal feedback consisting of full post-edits or gold-standard translations. A simulation experiment on news data confirmed this theoretical finding. This research implies that well-known online structured predictors can be used for learning from weak feedback, without changes to the algorithms. What is open for change is the strength of feedback, allowing “light” post-edits from non-professionals. An application of this scenario to online learning for patent translation is a desideratum for future work.

We refer the reader to Sokolov et al. (2014) for more information on the theory and proof-of-concept experiments in the coactive learning framework.

Another attack at learning from weak feedback can be taken from the direction of *bandit learning*.<sup>4</sup> Learning from bandit feedback describes an online learning scenario, where on each of a sequence of rounds, a learning algorithm makes a prediction, and receives partial information in terms of feedback to a single predicted point. In difference to the full information

<sup>3</sup><http://www.cl.uni-heidelberg.de/pattr/>

<sup>4</sup>The name of this framework is inherited from a model where in each round a gambler pulls an arm of a different slot machine (one-armed bandit), with the goal of maximizing his reward relative to the maximal possible reward, without apriori knowledge of the optimal slot machine.

---

**Algorithm 1** Bandit Structured Prediction

---

- 1: Input: sequence of learning rates  $\gamma_t$
  - 2: Initialize  $w_0$
  - 3: **for**  $t = 0, \dots, T$  **do**
  - 4:   Observe  $x_t$
  - 5:   Calculate  $\mathbb{E}_{p_{w_t}(y'|x_t)}[\phi(x_t, y')]$
  - 6:   Sample  $\tilde{y}_t \sim p_{w_t}(y'|x_t)$
  - 7:   Obtain feedback  $\Delta(\tilde{y}_t)$
  - 8:   Update  $w_{t+1} = w_t - \gamma_t \Delta(\tilde{y}_t)(\phi(x_t, \tilde{y}_t) - \mathbb{E}_{p_{w_t}(y'|x_t)}[\phi(x_t, y')])$
- 

---

**Algorithm** Structured Dueling Bandits

---

- 1: Input:  $\gamma, \delta, w_0$
  - 2: **for**  $t = 0, \dots, T$  **do**
  - 3:   Observe  $x_t$
  - 4:   Sample unit vector  $u_t$  uniformly
  - 5:   Set  $w'_t = w_t + \delta u_t$
  - 6:   Compare  $\Delta(\hat{y}_{w_t}(x_t))$  to  $\Delta(\hat{y}_{w'_t}(x_t))$
  - 7:   **if**  $w'_t$  wins **then**
  - 8:      $w_{t+1} = w_t + \gamma u_t$
  - 9:   **else**
  - 10:     $w_{t+1} = w_t$
- 

supervised scenario, the learner does not know what the correct prediction looks like, nor what would have happened if it had predicted differently. Applied to SMT, this means that the learning algorithm only has access to a  $1 - \text{BLEU}$  loss evaluation of a predicted translation instead of obtaining a gold standard reference translation. Clearly, a one-shot user quality estimate of the predicted translation is easier and faster to obtain than light or full post-edits of predicted translations, or than reference translations generated from scratch. This framework can be seen as a first step towards personalized machine translation where a given large SMT system is adapted to a user solely by single-point user feedback to predicted structure.

The crucial steps of the response-based online learning protocol are instantiated in this approach as follows:

**Prediction:** Exploration/exploitation sampling of sentence translation.

**Response:** User feedback on loss value of sampled translation.

**Learning:** Stochastic update using unbiased estimate of gradient.

Sokolov et al. (2015) presented two algorithms for online structured prediction in SMT from bandit feedback that implement these ideas as follows. Algorithm 1 optimizes an expected  $1 - \text{BLEU}$  loss criterion (Och (2003), Smith and Eisner (2006), He and Deng (2012), Auli et al. (2014), Wuebker et al. (2015), *inter alia*) by performing simultaneous exploration/exploitation by sampling translations from a Gibbs model (line 6), and using the obtained user feedback (line 7) to perform an update in the negative direction of the instantaneous gradient (line 8).

The second algorithm extends Yue and Joachims (2009)'s dueling bandits algorithm to a Structured Dueling Bandits algorithm. It compares a current weight vector  $w_t$  with a neighboring point  $w'_t$  along a direction  $u_t$ , performing exploration (controlled by  $\delta$ , line 5) by probing random directions, and exploitation (controlled by  $\gamma$ , line 8) by taking a step into the winning direction. The comparison step in line 6 is adapted to structured prediction from the original algorithm by comparing the quality of  $w_t$  and  $w'_t$  via an evaluation of the losses  $\Delta(\hat{y}_{w_t}(x_t))$  and

full information		bandit information	
<b>in-domain SMT</b>	<b>out-domain SMT</b>	<b>DuelingBandit</b>	<b>BanditStruct</b>
0.2854	0.2579	0.2731 $\pm$ 0.001	0.2705 $\pm$ 0.001

Table 3: Corpus BLEU on test set for SMT domain adaptation from Europarl to NewsCommentary by  $k$ -best reranking.

$\Delta(\hat{y}_{w'_t}(x_t))$  of the structured arms corresponding to predicting the most probable translation under  $w_t$  and  $w'_t$ , respectively.

Sokolov et al. (2015) present an evaluation that follows the standard of simulating bandit feedback by evaluating task loss functions against gold standard structures without revealing them to the learner. Here the setup is a reranking approach to SMT domain adaptation where the  $k$ -best list of an out-of-domain model is re-ranked (without re-decoding) based on bandit feedback from in-domain data. This can also be seen as a simulation of personalized machine translation where a given large SMT system is adapted to a user solely by single-point user feedback to predicted structures.

The *out-of-domain* baseline SMT model is trained on 1.6 million parallel Europarl data and includes the English side of Europarl and *in-domain* NewsCommentary in the language model. The full-information *in-domain* SMT model gives an upper bound by MERT tuning the out-of-domain model on in-domain development data. Learning under bandit feedback started at the learned weights of the *out-of-domain* median model. It uses the parallel NewsCommentary data to simulate bandit feedback, by evaluating the sampled translation against the gold standard reference using as loss function  $\Delta$  a smoothed per-sentence  $1 - \text{BLEU}$  (by flooring zero  $n$ -gram counts to 0.01).

Table 3 shows the final results that were obtained by online-to-batch conversion where the model trained for 100 epochs on *in-domain* training data is evaluated on a separate *in-domain* test set. Results for Bandit Structured Prediction and Dueling Bandits are very close, however, both are significant improvements over the out-of-domain SMT model that even includes an in-domain language model. The range of possible improvements is given by the difference of the BLEU score of the in-domain model and the BLEU score of the out-of-domain model – nearly 3 BLEU points. Bandit learning can improve the out-of-domain baseline by 1.26 BLEU points (Bandit Structured Prediction) and by 1.52 BLEU points (Dueling Bandits). Clearly, a comparison between Bandit Structured Prediction and Dueling Bandits is skewed towards the latter approach that has access to two-point feedback instead of one-point feedback as in the former case. It has been shown that querying the loss function at two points leads to convergence results that closely resemble bounds for the full information case (Agarwal et al., 2010), however, such feedback is clearly twice as expensive and, depending on the application, might not be elicitable from users.

We refer the reader to Sokolov et al. (2015) for more information on data, model, and experiments.

## 5 Conclusion

We presented a comprehensive perspective on machine learning approaches that attempt to replace the full information supervised scenario by a setup in which supervision signals are extracted from responses of an extrinsic environment to system predictions. The discussed types of responses ranged from rankings deduced from performance of translations in extrinsic tasks such as cross-lingual retrieval, to improvements of structures by post-edits of professional translators, to partial feedback consisting of mere assessments of the quality of the predicted translation. We showed the efficacy of response-based learning in several simulation experi-

ments. Clearly, improvements over traditional full-information structured prediction cannot be expected from learning from such weaker types of feedback. Instead, the goal is to investigate learning situations in which full information is not available. Moreover, task-based feedback might be even preferable to independently created gold standard structures if the ultimate goal is improved performance of a translation-related extrinsic task.

In future work, we would like to apply response-based learning from weak feedback to real-life interactive scenarios. The new challenges of future work will be an investigation of response signals that can be elicited from humans efficiently and reliably, and still are informative enough for learning in SMT.

### Acknowledgements

This research was supported in part by DFG grant RI-2221/1-2 “Weakly Supervised Learning of Cross-Lingual Systems”.

### References

- Agarwal, A., Dekel, O., and Xiao, L. (2010). Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Conference on Learning Theory (COLT)*, Haifa, Israel.
- Auli, M., Galley, M., and Gao, J. (2014). Large-scale expected BLEU training of phrase-based reordering models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*.
- Bertoldi, N., Simianer, P., Cettolo, M., Wäschle, K., Federico, M., and Riezler, S. (2014). Online adaptation to post-edits for phrase-based statistical machine translation. *Machine Translation*, 29:309–339.
- Cettolo, M., Federico, M., and Bertoldi, N. (2010). Mining parallel fragments from comparable texts. In *IWSLT’10*.
- Chin, J., Heymans, M., Kojoukhov, A., Lin, J., and Tan, H. (2008). Cross-language information retrieval. Patent Application. US 2008/0288474 A1.
- Denkowski, M., Dyer, C., and Lavie, A. (2014). Learning from post-editing: Online model adaptation for statistical machine translation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL’14)*, Gothenburg, Sweden.
- Downey, D., Dumais, S., Liebling, D., and Horvitz, E. (2008). Understanding the relationship between searchers’ queries and information goals. In *Proceedings of the 17th ACM conference on Information and knowledge management (CIKM’08)*, Napa Valley, CA.
- Garcia, I. (2011). Translating by post-editing: is it the way forward? *Machine Translation*, 25:217–237.
- Green, S., Heer, J., and Manning, C. D. (2013). The efficacy of human post-editing for language translation. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, Paris, France.
- Green, S., Wang, S. I., Chuang, J., Heer, J., Schuster, S., and Manning, C. D. (2014). Human effort and machine learnability in computer aided translation. In *Proceedings the onference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, Doha, Qatar.

- He, X. and Deng, L. (2012). Maximum expected BLEU training of phrase and lexicon translation models. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL'12)*, Jeju Island, Korea.
- Hieber, F. and Riezler, S. (2015). Bag-of-words forced decoding for cross-lingual information retrieval. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies (NAACL-HLT)*, Denver, CO.
- Koehn, P. (2009). A process study of computer-aided translation. *Machine Translation*, 23:241–263.
- Nikoulina, V., Kovachev, B., Lagos, N., and Monz, C. (2012). Adaptation of statistical machine translation model for cross-lingual information retrieval in a service context. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL'12)*, Avignon, France.
- Och, F. J. (2003). Minimum error rate training in statistical machine translation. In *Proceedings of the Human Language Technology Conference and the 3rd Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL'03)*, Edmonton, Canada.
- Smith, D. A. and Eisner, J. (2006). Minimum risk annealing for training log-linear models. In *Proceedings of the joint conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (COLING-ACL'06)*, Sydney, Australia.
- Sokolov, A., Hieber, F., and Riezler, S. (2014). Learning to translate queries for CLIR. In *Proceedings of the ACM SIGIR Conference*, Gold Coast, Australia.
- Sokolov, A., Riezler, S., and Urvoy, T. (2015). Bandit structured prediction for learning from user feedback in statistical machine translation. In *Proceedings of MT Summit XV*, Miami, FL.
- Ture, F., Lin, J., and Oard, D. W. (2012a). Combining statistical translation techniques for cross-language information retrieval. In *Proceedings of the International Conference on Computational Linguistics (COLING 2012)*.
- Ture, F., Lin, J., and Oard, D. W. (2012b). Looking inside the box: Context-sensitive translation for cross-language information retrieval. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2012)*, Portland, OR.
- Wäschle, K., Simianer, P., Bertoldi, N., Riezler, S., and Federico, M. (2013). Generative and discriminative methods for online adaptation in SMT. In *Proceedings of the Machine Translation Summit XIV*, Nice, France.
- Wuebker, J., Muehr, S., and Stephan Peitz, P. L., and Ney, H. (2015). A comparison of update strategies for large-scale maximum expected bleu training. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies (NAACL-HLT 2015)*, Denver, CO.
- Yue, Y. and Joachims, T. (2009). Interactively optimizing information retrieval systems as a dueling bandits problem. In *International Conference on Machine Learning (ICML)*, Montreal, Canada.