Coactive Learning for Interactive Machine Translation

Artem Sokolov
Computational Linguistics
Heidelberg University
69120 Heidelberg, Germany
sokolov@cl.uni-heidelberg.de

Stefan Riezler
Computational Linguistics & IWR
Heidelberg University
69120 Heidelberg, Germany
riezler@cl.uni-heidelberg.de

Shay B. Cohen
University of Edinburgh
Edinburgh EH8 9LE, UK
scohen@inf.ed.ac.uk

Abstract

Coactive learning describes the interaction between an online structured learner and a human user who corrects the learner by responding with weak feedback, that is, with an improved, but not necessarily optimal, structure. We apply this framework to discriminative learning in interactive machine translation. We present a generalization to latent variable models and give regret and generalization bounds for online learning with a feedback-based latent perceptron. We show experimentally that learning from weak feedback in machine translation leads to convergence in regret and translation error.

1 Introduction

In recent years, there has been a growing interest in "interactive machine translation" – a translation process in which outputs of a statistical machine translation (SMT) systems are being refined by humans, and used as training data in another machine translation learning phase. This gives rise to the question: to what degree should humans put an effort into giving exact correct outputs for the translation system?

In this paper, we pose this question and investigate it in the model of coactive learning [Shivaswamy and Joachims, 2012] – learning from weak feedback, in which non-exact inputs are available to the learning algorithm. The framework of coactive learning describes the interaction between a structured learning system and a human user where both have the same goal of providing results of maximum utility. The interaction follows an online learning protocol, where at each round t, the learner predicts a structured object y_t for an input x_t, and the user corrects the learner by responding with an improved, but not necessarily optimal, object $\bar{y}_t$ with respect to a utility function U. The key asset of coactive learning is the ability of the learner to converge to predictions that are close to optimal structures y_t^*, although the utility function is unknown to the learner, and only weak feedback in form of slightly improved structures $\bar{y}_t$ is seen in training.

The goal of this paper is to present a generalization of the framework of Shivaswamy and Joachims [2012] to latent variable models that are suitable for SMT, and give regret and generalization bounds for a feedback-based latent perceptron algorithm. Similar to the fully observable case, we show convergence at a rate of $O\big(\frac{1}{\sqrt{T}}\big)$, with possible improvements by using re-scaling in the algorithm. Furthermore, we present a proof-of-concept experiment that confirms our theoretical analysis by showing convergence in regret for learning from weak and strong feedback.

2 Related Work

Online learning from post-edits has mostly been confined to "simulated post-editing" where independently created human reference translations, or post-edits on the output from similar SMT systems, are used as for online learning (Cesa-Bianchi et al. [2008], López-Salcedo et al. [2012], Martínez-Gómez et al. [2012], *inter alia*). Most approaches rely on hidden derivation variables, thus they should be formalized as latent variable algorithms. To our knowledge, the aspect of learning from weak feedback has not been investigated so far in this area.

3 Feedback-based Latent Perceptron

Let $\mathcal{X}$ denote a set of input examples, e.g., sentences, and let $\mathcal{Y}(x)$ denote a set of structured outputs for $x \in \mathcal{X}$, e.g., translations. We define $\mathcal{Y} = \cup_x \mathcal{Y}(x)$. Furthermore, by $\mathcal{H}(x, y)$ we denote a set of possible hidden derivations for a structured output $y \in \mathcal{Y}(x)$, e.g., for phrase-based SMT, the hidden derivation is determined by a phrase segmentation and a phrase alignment between source and tar-

Appearing in Proceedings of the 4^{th} Workshop on Machine Learning for Interactive Systems (MLIS) 2015, Lille, France. JMLR: W&CP volume 40.

Algorithm 1 Feedback-based Latent Perceptron

1: Initialize $w \leftarrow 0$
2: **for** $t = 1, \ldots, T$ **do**
3: Observe x_t
4: $(y_t, h_t) \leftarrow \arg\max_{(y,h)} w_t^\top \phi(x_t, y, h)$
5: Obtain weak feedback $\bar{y}_t$
6: **if** $y_t \neq \bar{y}_t$ **then**
7: $\bar{h}_t \leftarrow \arg\max_h w_t^\top \phi(x_t, \bar{y}_t, h)$
8: $w_{t+1} \leftarrow w_t + \Delta_{\bar{h}_t, h_t}\big(\phi(x_t, \bar{y}_t, \bar{h}_t) - \phi(x_t, y_t, h_t)\big)$

get sentences. Every hidden derivation $h \in \mathcal{H}(x, y)$ deterministically identifies an output $y \in \mathcal{Y}(x)$. We define $\mathcal{H} = \cup_{x,y}\mathcal{H}(x, y)$. Let $\phi\colon \mathcal{X} \times \mathcal{Y} \times \mathcal{H} \to \mathbb{R}^d$ denote a feature function that maps a triplet (x, y, h) to a d-dimensional vector. For phrase-based SMT, we use 14 features, defined by phrase translation probabilities, language model probability, distance-based and lexicalized reordering probabilities, and word and phrase penalty. We assume that the feature function has a bounded radius, i.e. that $\|\phi(x, y, h)\| \leq R$ for all x, y, h. By $\Delta_{h,h'}$ we denote a distance function that is defined for any $h, h' \in \mathcal{H}$, and is used to scale the step size of updates during learning. In our experiments, we use the ordinary Euclidean distance between the feature vectors of derivations. We assume a linear model with fixed parameters w_* such that each input example is mapped to its correct derivation and structured output by using $(y^*, h^*) = \arg\max_{y \in \mathcal{Y}(x), h \in \mathcal{H}(x,y)} {w_*}^\top \phi(x, y, h)$.

Algorithm 1 is called "Feedback-based Latent Perceptron" to stress the fact that it only uses weak feedback to its predictions for learning, but does not necessarily observe optimal structures as in the full information case [Sun et al., 2013]. Learning from full information can be recovered by setting the informativeness parameter α to 1 in equation (2) below, in which case the feedback structure $\bar{y}_t$ equals the optimal structure y_t^*. Note that the maximization in line 7 can be replaced by a minimization or a random choice without loss of generality. In our theoretical exposition, we assume that $\bar{y}_t$ is reachable in the search space of possible outputs, that is, $\bar{y}_t \in \mathcal{Y}(x_t)$.

The key in the theoretical analysis in Shivaswamy and Joachims [2012] is the notion of a linear utility function $U_h(x, y) = {w_*}^\top \phi(x, y, h)$ determined by parameter vector w_*, that is unknown to the learner. Upon a system prediction, the user approximately maximizes utility, and returns an improved object $\bar{y}_t$ that has higher utility than the predicted structure y_t such that $U(x_t, \bar{y}_t) > U(x_t, y_t)$, where for given $x \in \mathcal{X}$, $y \in \mathcal{Y}(x)$, and $h^* = \arg\max_{h \in \mathcal{H}(x,y)} U_h(x, y)$, we define $U(x, y) = U_{h^*}(x, y)$ and drop the subscript unless $h \neq h^*$. Importantly, the feedback is typically not the optimal structure $y_t^* = \arg\max_{y \in \mathcal{Y}(x_t)} U(x_t, y)$. While not receiving optimal structures in training, the learning goal is to predict objects with utility close to optimal structures y_t^*. The regret that is suffered by the algorithm when predicting object y_t instead y_t^* is

$$\mathrm{REG}_T = \frac{1}{T}\sum_{t=1}^{T}\big(U(x_t, y_t^*) - U(x_t, y_t)\big). \qquad (1)$$

To quantify the amount of information in the weak feedback, Shivaswamy and Joachims [2012] define a notion of α-*informative* feedback, which we generalize as follows for the case of latent derivations. We assume that there exists a derivation $\bar{h}_t$ for the feedback structure $\bar{y}_t$, such that for all predictions y_t, the (re-scaled) utility of the weak feedback $\bar{y}_t$ is higher than the (re-scaled) utility of the prediction y_t by a fraction α of the maximum possible utility range (under the given utility model). Thus $\forall t, \exists \bar{h}_t, \forall h$ and for $\alpha \in (0, 1]$:

$$\begin{aligned}\big(U_{\bar{h}_t}(x_t, \bar{y}_t) - U_h(x_t, y_t)\big) \times \Delta_{\bar{h}_t, h} \\ \geq \alpha\big(U(x_t, y_t^*) - U(x_t, y_t)\big) - \xi_t, \qquad (2)\end{aligned}$$

where $\xi_t \geq 0$ are slack variables allowing for violations of (2) for given α. For slack $\xi_t = 0$, user feedback is called *strictly α-informative*.

4 Theoretical Analysis

A central theoretical result in learning from weak feedback is an analysis that shows that Algorithm 1 minimizes an upper bound on the average regret (1), despite the fact that optimal structures are not used in learning:

Theorem 1. *Let $D_T = \sum_{t=1}^{T} \Delta^2_{\bar{h}_t, h_t}$. Then the average regret of the feedback-based latent perceptron can be upper bounded for any $\alpha \in (0, 1]$, for any $w_* \in \mathbb{R}^d$:*

$$\mathrm{REG}_T \leq \frac{1}{\alpha T}\sum_{t=1}^{T}\xi_t + \frac{2R\|w_*\|}{\alpha}\frac{\sqrt{D_T}}{T}.$$

A proof for Theorem 1 is similar to the proof of Shivaswamy and Joachims [2012] and the original mistake bound for the perceptron of Novikoff [1962].[1] The theorem can be interpreted as follows: we expect lower average regret for higher values of α; due to the dominant term T, regret will approach the minimum of the accumulated slack (in case feedback structures violate equation (2)) or 0 (in case of strictly α-informative feedback). The main difference between the above result and the result of Shivaswamy and Joachims [2012] is the term D_T following from the re-scaled distance of latent derivations. Their analysis is agnostic of latent derivations, and can be recovered by setting this scaling factor to 1. This yields $D_T = T$, and thus recovers the main factor $\frac{\sqrt{D_T}}{T} = \frac{1}{\sqrt{T}}$ in their regret bound. In our algorithm, penalizing large distances of derivations can

[1] A short proof of the theorem is provided in the appendix.

	strict ($\xi_t = 0$)	slack ($\xi_t > 0$)
# datapoints	5,725	1,155
$\text{TER}(\bar{y}_t) < \text{TER}(y_t)$	52.17%	32.55%
$\text{TER}(\bar{y}_t) = \text{TER}(y_t)$	23.95%	20.52%
$\text{TER}(\bar{y}_t) > \text{TER}(y_t)$	23.88%	46.93%

Table 1: Improved utility vs. improved TER distance to human post-edits for α-informative feedback $\bar{y}_t$ compared to prediction y_t using default weights at $\alpha = 0.1$.

help to move derivations h_t closer to $\bar{h}_t$, therefore decreasing D_T as learning proceeds. Thus in case $D_T < T$, our bound is better than the original bound of Shivaswamy and Joachims [2012] for a perceptron without re-scaling. As we will show experimentally, re-scaling leads to a faster convergence in practice.

Furthermore, we can obtain a generalization bound for the case of online learning on a sequence of random examples, based on generalization bounds for expected average regret as given by Cesa-Bianchi et al. [2004]. Let probabilities $\mathbb{P}$ and expectations $\mathbb{E}$ be defined with respect to the fixed unknown underlying distribution according to which all examples are drawn. Furthermore, we bound our loss function $\ell_t = U(x_t, y_t^*) - U(x_t, y_t)$ to $[0, 1]$ by adding a normalization factor $2R||w_*||$ s.t. $\text{REG}_T = \frac{1}{T}\sum_{t=1}^{T} \ell_t$. Plugging the bound on REG_T of Theorem 1 directly into Proposition 1 of Cesa-Bianchi et al. [2004] gives the following theorem:

Theorem 2. *Let $0 < \delta < 1$, and let $x_1, \ldots, x_T$ be a sequence of examples that Algorithm 1 observes. Then with probability at least $1 - \delta$,*

$$\mathbb{E}[\text{REG}_T] \leq \frac{1}{\alpha T}\sum_{t=1}^{T}\xi_t + \frac{2R\|w_*\|}{\alpha}\frac{\sqrt{D_T}}{T} + 2||w_*||R\sqrt{\frac{2}{T}\ln\frac{1}{\delta}}.$$

5 Experiments

In this experiment, we apply Algorithm 1 to user feedback of varying utility grade. The goal of this experiment is to confirm our theoretical analysis by showing convergence in regret for learning from weak and strong feedback. We select feedback of varying grade by directly inspecting the optimal w_*. This setting can be thought of as an idealized scenario where a user picks translations from the n-best list that are considered improvements under the optimal w_*. However, the experiment also has a realistic background since we show that α-informative feedback corresponds to improvements under standard evaluation metrics such as lowercased and tokenized TER [Snover et al., 2006], and that learning from weak and strong feedback leads to convergence in TER on test data.

We used the LIG corpus[2] which consists of 10,881 tuples of French-English post-edits [Potet et al., 2012]. The corpus is a subset of the news-commentary dataset provided at WMT[3] and contains input French sentences, MT outputs, post-edited outputs and English references. To prepare SMT outputs for post-editing, the creators of the corpus used their own WMT10 system [Potet et al., 2010], based on the Moses phrase-based decoder[4] [Koehn et al., 2007] with dense features. We replicated a similar Moses system using the same monolingual and parallel data: a 5-gram language model was estimated with the KenLM toolkit [Heafield, 2011] on `news.en` data (48.65M sentences, 1.13B tokens), pre-processed with the tools from the `cdec`[5] toolkit. Parallel data (`europarl+news-comm`, 1.64M sentences) were similarly pre-processed and aligned with `fast_align` [Dyer et al., 2013]. In all experiments, training is started with the Moses default weights. The size of the n-best list, where used, was set to 1,000. Irrespective of the use of re-scaling in perceptron training, a constant learning rate of 10^{-5} was used for learning from simulated feedback, and 10^{-4} for learning from user post-edits. The post-edit data from the LIG corpus were randomly split into 3 subsets: PE-train (6,881 sentences), PE-dev, and PE-test (2,000 sentences each). PE-test was held out for testing the algorithms' progress on unseen data. PE-dev was used to obtain w_* to define the utility model. This was done by MERT optimization [Och, 2003] towards post-edits under the TER target metric. PE-train was used for our online learning experiments. The feedback data in this experiment were generated by searching the n-best list for translations that are α-informative at $\alpha \in \{0.1, 0.5, 1.0\}$ (with possible non-zero slack). This is achieved by scanning the n-best list output for every input x_t and returning the first $\bar{y}_t \neq y_t$ that satisfies Equation (2).[6]

In order to verify that our notion of graded utility corresponds to a realistic concept of graded translation quality, we compared improvements in utility to improved TER distance to human post-edits. Table 1 shows that for predictions under default weights, we obtain strictly α-informative (for $\alpha = 0.1$) feedback for 5,725 out of 6,881 datapoints in PE-train. These feedback structures improve utility per definition, and they also yield better TER distance to post-edits in the majority of cases. A non-negative slack has to be used in 1,155 datapoints. Here the majority

[2] `http://www-clips.imag.fr/geod/User/marion.potet/index.php?page=download`

[3] `http://www.statmt.org/wmt10/translation-task.html`

[4] `http://www.statmt.org/moses`

[5] `http://www.cdec-decoder.org/`

[6] Note that feedback provided in this way might be stronger than required at a particular value of α since for all $\beta \geq \alpha$, strictly β-informative feedback is also strictly α-informative. On the other hand, because of the limited size of the n-best list, we cannot assume strictly α-informative user feedback with zero slack ξ_t. In experiments where updates are only done if feedback is strictly α-informative we found similar convergence behavior.

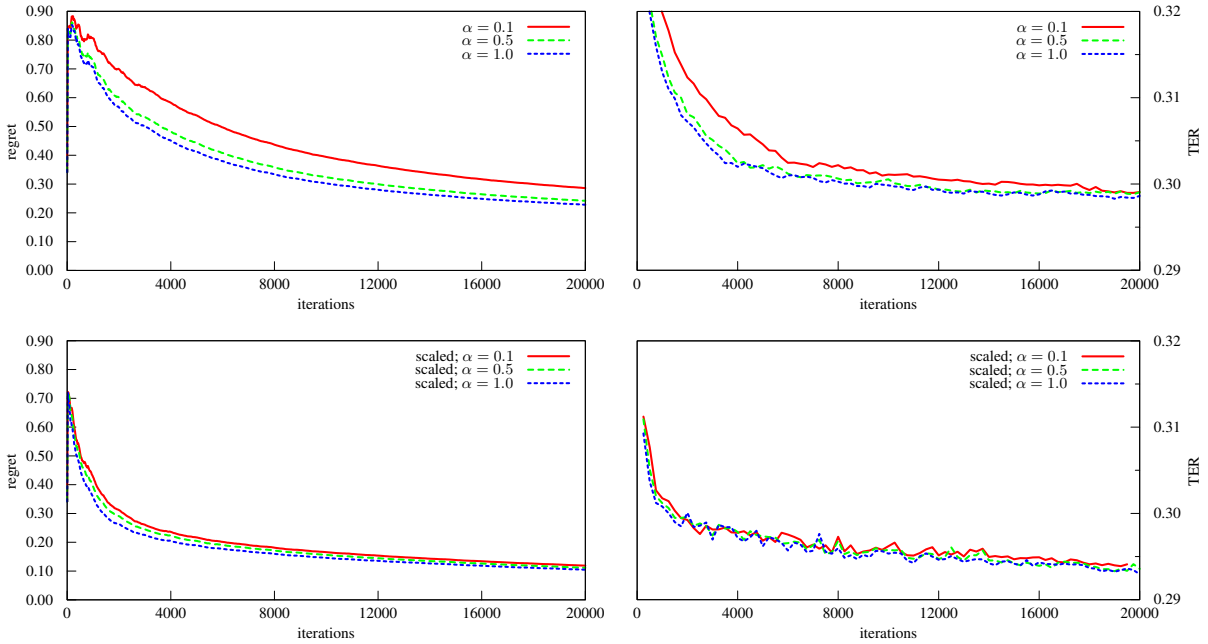

Figure 1: Regret and TER vs. iterations for α-informative feedback ranging from weak ($\alpha = 0.1$) to strong ($\alpha = 1.0$) informativeness, with (lower part) and without re-scaling (upper part).

of feedback structures do not improve TER distance.

Convergence results for different learning scenarios are shown in Figure 1. The left upper part of Figure 1 shows average utility regret against iterations for a setup without re-scaling, i.e., setting $\Delta_{\bar{h},h} = 1$ in the definition of α-informative feedback (Equation (2)) and in the update of Algorithm 1 (line 8). As predicted by our regret analysis, higher α leads to faster convergence, but all three curves converge towards a minimal regret[7]. Also, the difference between the curves for $\alpha = 0.1$ and $\alpha = 1.0$ is much smaller than a factor of ten. As expected from the correspondence of α-informative feedback to improvements in TER, similar relations are obtained when plotting TER scores on test data for training from weak feedback at different utility grades. This is shown in the right upper part of Figure 1. The left lower part of Figure 1 shows average utility regret plotted against iterations for a setup that uses re-scaling. We define $\Delta_{\bar{h}_t,h}$ by the ℓ_2-distance between the feature vectors $\phi(x_t, \bar{y}_t, \bar{h}_t)$ of the derivation of the feedback structure and the feature vector $\phi(x_t, y_t, h_t)$ of the derivation of the predicted structure. We see that convergence in regret is faster for re-scaling. Furthermore, as shown in the right lower part of Figure 1, TER is decreased on test data as well at a faster rate.

[7]We stopped learning at a regret value of about 0.1 .

6 Discussion

We presented an extension of Shivaswamy and Joachims [2012]'s framework of coactive learning to interactive SMT where a human user corrects an online structured learning system by post-editing a predicted translation.

In an experiment on learning from simulated weak and strong feedback, we confirmed convergence for learning from weak feedback, with faster convergence for stronger feedback and for rescaling the learning rate. This experiment can be thought of as an idealized scenario in which the user has access to the optimal utility function. A possible extension of this experiment would be to investigate a scenario where users pick translations from the n-best list that they consider improvements over the prediction.

In future work we would like to show that for the area of interactive SMT, "light" post-edits might be preferable over "full" post-edits because they are better reachable, easier elicitable, and yet provide a strong enough signal for learning.

Acknowledgments

This research was supported in part by DFG grant RI-2221/2-1 "Grounding Statistical Machine Translation in Perception and Action".

Appendix: Proof of Theorem 1

Proof. First we bound $w_{T+1}^\top w_{T+1}$ from above:

$$\begin{aligned} w_{T+1}^\top w_{T+1} &= w_T^\top w_T \\ &+ 2w_T^\top \big(\phi(x_T, \bar{y}_T, \bar{h}_T) - \phi(x_T, y_T, h_T)\big)\Delta_{\bar{h}_T, h_T} \\ &+ \big(\phi(x_T, \bar{y}_T, \bar{h}_T) - \phi(x_T, y_T, h_T)\big)^\top \Delta_{\bar{h}_T, h_T} \\ &\quad \big(\phi(x_T, \bar{y}_T, \bar{h}_T) - \phi(x_T, y_T, h_T)\big)\Delta_{\bar{h}_T, h_T} \\ &\leq w_T^\top w_T + 4R^2\Delta_{\bar{h}_T, h_T}^2 \leq 4R^2 D_T. \end{aligned} \tag{3}$$

The first equality uses the update rule from Algorithm 1. The second uses the fact that $w_T^\top(\phi(x_T, \bar{y}_T, \bar{h}_T) - \phi(x_T, y_T, h_T)) \leq 0$ by definition of (y_T, h_T) in Algorithm 1. By assumption $\|\phi(x, y, h)\| \leq R, \forall x, y, h$ and by the triangle inequality, $\|\phi(x, y, h) - \phi(x, y', h')\| \leq \|\phi(x, y, h)\| + \|\phi(x, y', h')\| \leq 2R$. Finally, $D_T = \sum_{t=1}^T \Delta_{\bar{h}_t, h_t}^2$ by definition, and the last inequality follows by induction.

The connection to average regret is as follows:

$$\begin{aligned} w_{T+1}^\top w_* &= w_T^\top w_* \\ &+ \Delta_{\bar{h}_T, h_T}\big(\phi(x_T, \bar{y}_T, \bar{h}_T)) - \phi(x_T, y_T, h_T)\big)^\top w_* \\ &= \sum_{t=1}^T \Delta_{\bar{h}_t, h_t}\big(\phi(x_t, \bar{y}_t, \bar{h}_t) - \phi(x_t, y_t, h_t)\big)^\top w_* \\ &= \sum_{t=1}^T \Delta_{\bar{h}_t, h_t}\big(U_{\bar{h}_t}(x_t, \bar{y}_t) - U_{h_t}(x_t, y_t)\big). \end{aligned} \tag{4}$$

The first equality again uses the update rule from Algorithm 1. The second follows by induction. The last equality applies the definition of utility. Next we upper bound the utility difference:

$$\begin{aligned} &\sum_{t=1}^T \Delta_{\bar{h}_t, h_t}\big(U_{\bar{h}_t}(x_t, \bar{y}_t) - U_{h_t}(x_t, y_t)\big) \\ &\qquad \leq \|w_*\|\|w_{T+1}\| \leq \|w_*\| 2R\sqrt{D_T}. \end{aligned} \tag{5}$$

The first inequality follows from applying the Cauchy-Schwartz inequality $w_{T+1}^\top w_* \leq \|w_*\|\|w_{T+1}\|$ to Equation (4). The seond follows from applying Equation (3) to $\|w_{T+1}\| = \sqrt{w_{T+1}^\top w_{T+1}}$. The final result is obtained simply by lower bounding Equation (5) using the assumption in Equation (2).

$$\begin{aligned} \|w_*\| 2R\sqrt{D_T} &\geq \sum_{t=1}^T \Delta_{\bar{h}_t, h_t}\big(U_{\bar{h}_t}(x_t, \bar{y}_t) - U_{h_t}(x_t, y_t)\big) \\ &\geq \alpha \sum_{t=1}^T \big(U(x_t, y_t^*) - U(x_t, y_t)\big) - \sum_{t=1}^T \xi_t \\ &= \alpha\, T\, \mathrm{REG}_T - \sum_{t=1}^T \xi_t. \end{aligned}$$

□

References

Nicolo Cesa-Bianchi, Alex Conconi, and Claudio Gentile. On the generalization ablility of on-line learning algorithms. *IEEE Transactions on Information Theory*, 50 (9):2050–2057, 2004.

Nicolò Cesa-Bianchi, Gabriele Reverberi, and Sandor Szedmak. Online learning algorithms for computer-assisted translation. Technical report, SMART (`www.smart-project.eu`), 2008.

Chris Dyer, Victor Chahuneau, and Noah A. Smith. A simple, fast, and effective reparameterization of IBM model 2. In *HLT-NAACL*, Atlanta, GA, 2013.

Kenneth Heafield. KenLM: faster and smaller language model queries. In *WMT*, Edinburgh, Scotland, UK, 2011.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Birch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. Moses: Open source toolkit for statistical machine translation. In *ACL Demo and Poster Sessions*, Prague, Czech Republic, 2007.

Francisco-Javier López-Salcedo, Germán Sanchis-Trilles, and Francisco Casacuberta. Online learning of log-linear weights in interactive machine translation. In *IberSpeech*, Madrid, Spain, 2012.

Pascual Martínez-Gómez, Germán Sanchis-Trilles, and Francisco Casacuberta. Online adaptation strategies for statistical machine translation in post-editing scenarios. *Pattern Recognition*, 45(9):3193–3202, 2012.

Albert B.J. Novikoff. On convergence proofs on perceptrons. *Symposium on the Mathematical Theory of Automata*, 12:615–622, 1962.

Franz Josef Och. Minimum error rate training in statistical machine translation. In *HLT-NAACL*, Edmonton, Canada, 2003.

Marion Potet, Laurent Besacier, and Hervé Blanchon. The LIG machine translation system for WMT 2010. In *WMT*, Upsala, Sweden, 2010.

Marion Potet, Emanuelle Esperança-Rodier, Laurent Besacier, and Hervé Blanchon. Collection of a large database of French-English SMT output corrections. In *LREC*, Istanbul, Turkey, 2012.

Pannaga Shivaswamy and Thorsten Joachims. Online structured prediction via coactive learning. In *ICML*, Scotland, UK, 2012.

Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A study of translation edit rate with targeted human annotation. In *AMTA*, Cambridge, MA, 2006.

Xu Sun, Takuya Matsuzaki, and Wenjie Li. Latent structured perceptrons for large scale learning with hidden information. *IEEE Transactions on Knowledge and Data Engineering*, 25(9):2064–2075, 2013.