Iterative, MT-based Sentence Alignment of Parallel Texts

Rico Sennrich and **Martin Volk**
Insitute of Computational Linguistics
University of Zurich
Binzmühlestr. 14
CH-8050 Zürich
{sennrich,volk}@cl.uzh.ch

Abstract

Recent research has shown that MT-based sentence alignment is a robust approach for noisy parallel texts. However, using Machine Translation for sentence alignment causes a chicken-and-egg problem: to train a corpus-based MT system, we need sentence-aligned data, and MT-based sentence alignment depends on an MT system. We describe a bootstrapping approach to sentence alignment that resolves this circular dependency by computing an initial alignment with length-based methods. Our evaluation shows that iterative MT-based sentence alignment significantly outperforms widespread alignment approaches on our evaluation set, without requiring any linguistic resources other than the to-be-aligned bitext.

1 Introduction

Given a parallel text, i.e. the same text in two (or more) languages, aligning the different language versions on a sentence level is a necessary first step for corpus-based machine translation (e.g. statistical MT (SMT) or example-based MT), but also for building translation memories from existing parallel texts or other forms of multilingual analysis. Some parallel texts can be aligned with comparative ease. Parliamentary Proceedings such as Europarl or the Canadian Hansards, which are frequently used for MT, provide markup information to identify the different speakers; such markup information provides useful anchor points for an alignment, and allows for accuracies above 95% (Gale and Church, 1993).

However, sentence alignment is significantly harder for other texts, as will be illustrated in section 3. Since SMT systems depend on relevant training data for their performance, it is not sufficient to only use easily accessible and alignable texts as training material for SMT systems; ideally, SMT systems should be trained on texts that are similar to those one wishes to translate. This warrants continued research on more robust sentence alignment algorithms.

Bleualign is a sentence alignment algorithm that, instead of computing an alignment between the source and target text directly, bases its alignment search on an MT translation of the source text. It has been shown that Bleualign can robustly align texts for which other algorithms perform poorly (Sennrich and Volk, 2010). The quality of sentence alignment has an effect on the performance of SMT systems (Lambert et al., 2010), but high quality is also desirable for other purposes, e.g. when building a translation memory from a text corpus. The main disadvantage of an MT-based algorithm is that it requires an existing MT system. For resource-poor language pairs, this requirement makes the algorithm unattractive.

We have investigated the bootstrapping of MT-based sentence alignment with an MT system trained on the to-be-aligned corpus. For this first MT system, a length-based sentence alignment algorithm is used which requires no linguistic resources. Such an iterative approach can supersede the dependence of the algorithm on existing MT systems.

(Sennrich and Volk, 2010) have demonstrated that the sentence alignment quality of their MT-based algorithm depends on the quality of the MT system. If we can produce a superior MT system using Bleualign, it is worthwhile to test if the resulting MT system can in turn be used for an even

Bolette Sandford Pedersen, Gunta Nešpore and Inguna Skadiņa (Eds.)
NODALIDA 2011 Conference Proceedings, pp. 175–182

better sentence alignment.

2 Related Work

The first sentence alignment algorithms by (Brown et al., 1991) and (Gale and Church, 1993) are based on a length-comparison between source and target text and work without language-specific information.[1] A second strand of sentence alignment algorithms work with lexical correspondences. This is either done on the basis of correspondence rules (Simard et al., 1993), with external dictionaries (Varga et al., 2005), or using a translation model trained on the parallel text itself (Moore, 2002; Varga et al., 2005). The latter requires a preliminary sentence alignment of the parallel text, usually performed with a length-based algorithm. After this first pass, a translation model can be trained (e.g. an IBM Model 1 in the case of (Moore, 2002), a dictionary-based translation model in (Varga et al., 2005)), which is then used for the alignment in a second pass.

(Sennrich and Volk, 2010) describe an alignment algorithm based on the automatic translation of one language portion of the parallel text. They use existing MT systems to translate the to-be-aligned parallel text, then try to find an alignment between the translated source text and the target text that maximizes the BLEU score. Since sentence alignment is required to build an SMT system, being dependent on existing MT systems for sentence alignment causes a chicken-and-egg problem.

Circular dependencies as the one relating to MT-based sentence alignment are a well-known problem, for instance for word alignment. Word translation probabilities can only be estimated from a word-aligned parallel text, and to word-align the parallel text, we need a translation model. Brown et al. (Brown et al., 1993) use an iterative Expectation-Maximization algorithm for word alignment in the still widely-used IBM models.

In this paper, we investigate if the algorithm lends itself to an iterative approach similar in spirit to the one by (Brown et al., 1993), in order to avoid the dependency on pre-existing MT systems for sentence alignment[2], while obtaining equal or better results.

3 The Parallel Text

We conduct our experiments on the parallel part of the Text+Berg corpus, a collection of Alpine texts (Volk et al., 2010). As of now, the collection consists of the yearbooks of the Swiss Alpine Club from 1864 to 1995. Since 1957, the yearbook has been published in two parallel editions, German and French. This results in about 3 million tokens of parallel text which can be used for Statistical Machine Translation.

The Text+Berg corpus is characterized by its thematic homogeneity. The topic of most if not all texts are the mountains. However, there is a wide range of text types represented in the corpus. We will illustrate this with some examples from the 1975 yearbook. Most typical are reports on mountain expeditions: *Schreckhorn-Nordwand im Winter* (English: *Schreckhorn North Face in winter*). We also find poems (one called *Praise of Nature*, one on the alphorn), a historical account on the extermination and reintroduction of ibex in the Swiss Alps, and articles that capture current trends or innovations, such as *Segelflieger im Gebirge* (English: *Gliders in the Mountains*). Recurring articles include chronicles of Himalaya expeditions, and scientific reports on the periodic variations of the glaciers in the Swiss Alps.

The corpus poses interesting challenges for Machine Translation. The terminology used in the text is very specific and is translated badly by SMT systems trained on out-of-domain data. In a set of 1000 Text+Berg sentences, 11% of tokens, or 31.4% of types, are out-of-vocabulary items for a SMT system trained on the Europarl corpus. These unseen words can be roughly divided into the following categories: named entities (*Nadelhorn*; *Selbstsanft*), domain-specific vocabulary (*Pickel*, English: *pick-axe*; *Basislager*, English: *base camp*), Swiss spelling variations (*gross* instead of *groß*, English: *big*), and OCR errors (*iweimal* instead of *zweimal*, English: *twice*). Of course, we can expect a certain proportion of unseen words in any text, especially in German, where compounds and inflected word forms abound[3]. Domain-specific vocabulary is the most

[1] To be precise, Gale & Church's algorithm does contain a priori probabilities for deletions and insertions estimated from the Canadian Hansards (Gale and Church, 1993). However, these parameters are usually left untouched (Danielsson and Ridings, 1997).

[2] We use open source tools to build the SMT systems; we do not, however, use any training data other than the parallel text we wish to align.

[3] We measured 0.8% unseen tokens, 4.8% unseen types in

prevalent category in the list of unseen words, and is the strongest reason for adapting MT systems to new domains with in-domain training data – not only to reduce the number of unseen words, but also to learn domain-specific translations of polysemous terms. For instance, the German term *Führer* is usually translated into French as *dirigeant* in Europarl (English: *leader*), but as *guide* in Text+Berg.

Aligning the Text+Berg corpus on a sentence level is surprisingly difficult. The texts are aligned semi-automatically on an article level. Within each article, there are no reliable structural markers: the number of paragraphs is different for the two language versions; page breaks are at different places in the text. With an average article length of approximately 200 sentences (which is about 6 pages of text), the search space for possible alignments is significantly larger than for the small segments in Europarl which are delimited by commentary tags.

Additionally, the ratio of 1-to-1 aligned sentences (We will subsequently call any n-to-m alignment a *bead*) is very low in the articles which we manually aligned for evaluation purposes. Out of the 422 beads found in a manually aligned article, only 58.3% are 1-to-1 beads. This is a striking contrast to earlier publications on the topic of sentence alignment, which reported on texts with over 90% 1-to-1 beads (Manning and Schütze, 1999). In the hand-aligned article, 19.5% of the beads are 1-to-2 or 2-to-1, 9.7% deletions (0-to-1 or 1-to-0), and the remaining 12.6% beads of higher order (2-to-2, 1-to-many, many-to-1, many-to-many).

The two main reasons for the low number of 1-to-1 beads are the joining or splitting of sentences by the translators, and errors in the digitization of the corpus. The 1-to-4 bead shown in table 1 is an example of a German sentence being split up into several French ones. In the article we hand-aligned, the translator frequently splits or joins sentences in this way. We cannot claim that the article is representative of the whole corpus, however, hence we do not exactly know how pervasive this problem is[4]. Other 1-to-many alignments are artifacts of the digitization process, e.g. OCR, tokenization, or sentence boundary detection errors.

In summary, sentence alignment is considerably more difficult for the Text+Berg corpus than e.g. for Europarl, both because there are few anchor points, and because the number of 1-to-1 beads is low.

a Europarl test set, with a Europarl training set.

[4]In a independent hand-aligned set of 1000 sentences by various authors, we found 74% 1-to-1 alignments.

4 Iterative Sentence Alignment

Technically, iterative sentence alignment is simple, given freely available tools for SMT and sentence alignment. Each iteration consists of the following steps:

1. Sentence-align the parallel training corpus.
   - In the first iteration, use an implementation of the Gale & Church algorithm (or any other sentence alignment tool that does not require additional resources).
   - In all subsequent iterations:
     - Automatically translate the corpus using the SMT system trained in the last iteration.
     - Align the texts using Bleualign and this translation.
2. Train an SMT system on the sentence-aligned corpus.

The language model needs only be trained once; we use SRILM (Stolcke, 2002). The SMT system is built with GIZA++ (Och and Ney, 2003) and Moses (Koehn et al., 2007). The most time-consuming part of each iteration is typically the automatic translation of the training set.

In the remainder of this section, we will discuss how the alignment algorithm works, and what potential problems the iterative approach brings.

4.1 The Sentence Alignment Algorithm

The sentence alignment algorithm, first described in (Sennrich and Volk, 2010), is a two-pass approach. In the first pass, dynamic programming is used to find a set of 1-to-1 beads that maximizes BLEU score in the document without violating the monotonic order of sentence pairs. In the second pass, unaligned sentences are either added to beads found in the first pass (if warranted by increasing BLEU scores), aligned using a length-based algorithm (if possible without violating the monotonic order of sentence pairs), or discarded.

It was shown that the algorithm is very sensitive to the quality of the automatic translation (Sennrich and Volk, 2010). If no translation is provided, performance is actually worse than if the texts are aligned using the algorithm by Gale &

s_1	Aber hinter dem grossen Turm wird der Schnee grundlos, keiner von der ganzen Seilschaft hat sicheren Stand, die Spur wird zu einem tiefen Graben, der Mann an der Spitze wühlt sich 30, höchstens 40 Schritte aufwärts und tritt dann wortlos zur Seite, um dem nächsten Platz zu machen. [But behind the great tower, the snow becomes groundless; noone in the rope team has a secure footing. The track becomes a deep trench; the man in the vanguard climbs through the snow for 30, no more than 40 steps and then silently steps aside to make room for the next person.]
t_1	Mais au delà de la grosse tour, la neige est sans consistance;
t_2	aucun des membres de la cordée ne peut assurer solidement.
t_3	La trace devient une vraie tranchée;
t_4	le premier patauge péniblement pendant 30 , au maximum 40 pas, puis, sans un mot, tire de côté pour laisser place au suivant.

Table 1: Example of a 1-to-4 alignment. *s* is the German source text; *t* the French target text. English translation ours.

Church. This may happen if the BLEU-based first pass yields wrong beads, for instance if there are recurring names or dates.

4.2 Pruning

Let us consider the effect of misaligned sentence pairs. With the word alignment and phrase extraction algorithms that the Moses system uses, wrong phrase translations will be learned if sentences are misaligned. Such wrong phrase translations are normal in SMT, and usually not a big problem. For frequent phrases, every wrong phrase translation tends to be much rarer (and thus less probable) than correct ones. Rare phrases that are mistranslated are unlikely to occur again in the to-be-translated text.

Unfortunately, this last point does not hold true for an iterative approach where the training text is also the to-be-translated text. The type *AlbertEggler*, an artifact caused by OCR, only occurs once in the Text+Berg corpus. It is part of the sentence - *AlbertEggler :*, which is misaligned in the first sentence alignment pass to the sentence *1954 , Helmut Heuberger en géographie ;*. Consequently, the training algorithm estimates via Maximum Likelihood Estimation that the phrase *- AlbertEggler :* is translated to *1954 , Helmut Heuberger en géographie ;* with a probability of 1.[5] Hence, the sentence is mistranslated during the next iteration. The problem is that such mistranslations may cause the same alignment errors to be made in subsequent iterations.

In order to prevent random misalignments to be fossilized, we prune the translation model using the approach by (Johnson et al., 2007). The pruning is based on computing whether the co-occurrence frequency of phrase pairs in the translation model is statistically significant, or to be expected by chance. All phrase pairs whose significance value fall below a predefined threshold are discarded. We chose the significance threshold $\alpha + \epsilon$, which among others discards all phrase pairs that co-occur only once.[6]

5 Evaluation

For the evaluation of alignment quality, we manually aligned an article consisting of 468 and 554 sentences (German and French, respectively). This manual alignment serves as a gold standard to which the automatic sentence alignments will be compared. The alignment test set is a subset of the training set. This unusual choice was made because it mirrors the conditions of the iterative approach: the text that is to be translated serves as training set for the SMT system, which potentially causes errors (see section 4.2). To test whether pruning mitigates the problem, we will perform the evaluation both with and without pruning.

Because of the high proportion of 1-to-many alignments, we will use two different truth conditions, which are evaluated on a per-alignment basis. Under the strict truth condition, we demand an exact match between the gold alignment and the hypothesis. Under the lax condition, a hypothesis is true if there is an overlap with a gold alignment on both language sides. This means that a 2-to-2 alignment that is misrecognized as two 1-to-1

[5] The term *phrase* is used to denote arbitrary word sequences in SMT, without syntactic implications. In this case, the whole sentence is treated as a single phrase by the SMT system.

[6] with $\alpha = log(N)$ and ϵ an "appropriately small positive number" (Johnson et al., 2007).

Algorithm	Alignment based on	Alignment quality F_1 strict	F_1 lax	BLEU
G&C	-	0.2%	0.2%	15.54
Bleualign	Europarl	69.5%	94.4%	16.38

Table 2: Baseline scores: Sentence alignment quality and MT performance (with pruning). G&C: Gale & Church algorithm.

alignments will count as two false positives under the strict condition, but two true positives under the lax condition.

While sentence alignment may serve various goals, our main interest is using the aligned corpus for SMT, and obtaining better translation systems from better-aligned corpora. Hence, we measure translation performance of all SMT systems trained through BLEU (Papineni et al., 2002). The systems, built with SRILM (Stolcke, 2002), GIZA++ (Och and Ney, 2003) and Moses (Koehn et al., 2007), will be evaluated on a test set of 1000 sentences, held-out from training. The training set consists of 3 300 000 German and 3 740 000 French tokens, measured before sentence alignment.[7] We test translation performance in the direction DE–FR, and use a language model trained on 9 511 000 tokens of in-domain text. We did not perform Minimum Error Rate Training, which is typically the most time-intensive step of training an SMT system, in order to limit the computational cost of the iterative approach. Statistical significance is tested with paired bootstrap resampling (Koehn, 2004).

5.1 Results

We first establish baseline scores achieved by either using the Gale & Church algorithm or Bleualign with an out-of-domain MT system, shown in table 2. For a wider comparison of different sentence alignment algorithms, see (Lambert et al., 2010).

On the alignment test set, Gale & Church's algorithm fails almost entirely; only 1 out of 468 alignment hypotheses is correct. The reason for the bad performance of the Gale & Church algorithm in this evaluation is that errors tend to propagate, since misaligned sentences may cause neighbouring sentences to be misaligned as well. This is one of the reasons why anchor points – article boundaries in our case – are so important; they serve as boundaries to the alignment algorithm and stop the propagation of errors from one article to the next.

For the iterative approach, the results obtained by aligning the Text+Berg training corpus with Bleualign, based on a translation of the corpus with a SMT system trained on Europarl[8], serve as the baseline. We observe an F_1 score of 69.5% (strict condition) and 94.4% (lax condition) in the evaluation of alignment quality. With 16.38 BLEU points, it is significantly better in terms of MT performance than the system aligned with the Gale & Church algorithm (15.54 BLEU points).[9]

It might seem surprising that MT performance of the system that is based on the Gale & Church alignment is still acceptable, despite most alignments in the alignment test set being wrong. However, note that the MT quality evaluation is based on the entire Text+Berg corpus, whereas the alignment quality evaluation is based on a relatively small test set of about 500 setences; there are articles for which Gale & Church alignment performs better. In terms of how difficult the test set is to align, this evaluation is complementary to the one by (Sennrich and Volk, 2010), who evaluated alignment algorithms on a test set of seven shorter articles. Having a difficult-to-align test set is important for the second part of our evaluation; the high error rate of the Gale & Church algorithms for this test set allows us to observe whether and to what degree misalignments are self-reinforcing, as we outlined in section 4.2.

Table 3 shows SMT and alignment performance for each of 5 iterations. Table 4 does the same, but before re-translating the training text, the system is pruned according to (Johnson et al., 2007). Note that we are interested in the effects of pruning on the alignment of training data, not in the direct effect of pruning on SMT results. This is why even for the unpruned experiment (table 3), we show MT results both with and without pruning. The effect of pruning is especially strong in the first iteration (which is identical to the baseline Gale & Church system): pruning accounts for an increase in BLEU score from 13.72 to 15.54

[7] The final number used for training may vary, depending on the number of sentences discarded during alignment, and the number of sentence pairs filtered because of sentence length.

[8] Approximately 25 000 000 tokens per language for training the translation model, 47 000 000 French tokens for the language model.

[9] Note that the training corpus is the same for all experiments; only the alignment algorithm and the system used to translate the corpus change between experiments.

i	Algorithm	Alignment based on	Alignment quality		BLEU	
			F_1 strict	F_1 lax	no pruning	pruning
1	G&C	-	0.2%	0.2%	13.72	15.54
2	Bleualign	i 1	36.7%	63.8%	15.26	15.98
3	Bleualign	i 2	56.7%	86.1%	15.56	16.27
4	Bleualign	i 3	63.9%	92.8%	15.83	16.50
5	Bleualign	i 4	65.3%	94.0%	15.69	16.44

Table 3: Sentence alignment quality and MT performance after i iterations. For each alignment, the *unpruned* MT system from the previous iteration is used. G&C: Gale & Church algorithm.

BLEU points. In later iterations, the difference is between 0.7 and 0.8 BLEU points.

We can see that the alignment quality improves after each iteration in the experiment without pruning (table 3). However, after 4 iterations, it is still lower than in the baseline experiment with an SMT system trained on Europarl. The systems in the experiment with pruning reach a higher alignment quality, and reach it after fewer iterations. We cannot explain this difference away through the general quality increase through pruning. In the third and fourth iteration without pruning, MT performance on the held-out test set is higher than in the first iteration with pruning. Still, the pruned system leads to a higher alignment quality in the subsequent iteration. We conclude that a self-reinforcement of misalignments, as described in section 4.2, does indeed occur if we do not prune the SMT systems, and that pruning successfully combats this effect.

In this experiment, an iterative alignment (with pruning) only requires two iterations to reach a stable level both in alignment quality and SMT performance. SMT performance of the second iteration of the experiment with pruning is significantly better than both baselines, and significantly better than the fifth iteration without pruning. Compared to the baseline with the Europarl SMT system, the increase is relatively small, from 16.38 to 16.67 BLEU points. At least in this experiment, the main advantage of the iterative approach lies not in a performance increase, but in being independent from external MT systems.

5.2 Interpretation and Usage Recommendations

Having to translate the entire training corpus for sentence alignment is a costly requirement, even if the iterative algorithm does not rely on external MT systems. It is thus positive that, with our SMT tools and the well-known pruning approach by (Johnson et al., 2007), we reach the highest quality after just two iterations, meaning that the training corpus only needs to be translated once. Still, we do not recommend iterative sentence alignment with Bleualign for all purposes.

i	Algorithm	Alignment based on	Alignment quality		BLEU
			F_1 strict	F_1 lax	
1	G&C	-	0.2%	0.2%	15.54
2	Bleualign	i 1	76.1%	97.6%	16.67
3	Bleualign	i 2	76.7%	97.6%	16.60
4	Bleualign	i 3	76.1%	97.5%	16.64
5	Bleualign	i 4	76.4%	98.0%	16.52

Table 4: Sentence alignment and MT performance quality after i iterations (with pruning). G&C: Gale & Church algorithm.

Aspects worth considering for the choice of sentence alignment algorithm are:

1. The accuracy of computationally less expensive sentence alignment algorithms such as Gale & Church's on the parallel text. The lower their accuracy, the more promising it is to perform a sentence alignment with Bleualign.

2. The size of the parallel text. If the amount of parallel text is too small to train an adequate MT system with it, we recommend using Bleualign with a pre-existing MT system or a different alignment algorithm altogether. On the other hand, if the amount of parallel text is very large, this slows down the iterations considerably, both because of the large amount of text to be translated and the increase in training/decoding time resulting from more data. Using only a subsection of the parallel text to build the first, non-final SMT system will speed up the process.

3. The availability of language-specific resources. Whether the recommended resources are dictionaries (Varga et al., 2005), or MT systems (Sennrich and Volk, 2010), they might be unavailable or lacking in qual-

ity for a given alignment task. The iterative sentence alignment approach described in this paper is especially suitable for language pairs with few existing resources.

6 Conclusion

In (Sennrich and Volk, 2010), Bleualign was established as a well-performing sentence alignment tool given a sufficiently good existing MT system. In this paper, we show that a similar performance can be achieved without the use of language-specific resources other than the to-be-aligned parallel text. We do this by training an SMT system on the to-be-aligned text, using a length-based sentence alignment algorithm. This SMT system is then used to translate the source side of the parallel training corpus; on this translation, Bleualign bases its sentence alignment.

The biggest weakness of an iterative sentence alignment approach is that misaligned sentences lead to errors in the translation model, which tend to cause the same alignment errors in the next iteration. We show that pruning singleton phrase pairs improves the quality of iterative sentence alignment tremendously, leading to the best results after just two iterations.

Acknowledgments

This research was funded by the Swiss National Science Foundation under grant 105215_126999.

References

Peter F. Brown, Jennifer C. Lai, and Robert L. Mercer. 1991. Aligning sentences in parallel corpora. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, pages 169–176, Morristown, NJ, USA.

P.F. Brown, V.J. Della Pietra, S.A. Della Pietra, and R.L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2):263–311.

Pernilla Danielsson and Daniel Ridings. 1997. Practical presentation of a "vanilla" aligner. In *TELRI Workshop on Alignment and Exploitation of Texts*, Ljubljana. Institute Jozef Stefan.

William A. Gale and Kenneth W. Church. 1993. A program for aligning sentences in bilingual corpora. *Comput. Linguist.*, 19(1):75–102.

Howard Johnson, Joel Martin, George Foster, and Roland Kuhn. 2007. Improving translation quality by discarding most of the phrasetable. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 967–975, Prague, Czech Republic, June. Association for Computational Linguistics.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of ACL 2007*, pages 177–180, Prague, Czech Republic, June.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP 2004*, Barcelona, Spain.

Patrik Lambert, Sadaf Abdul-Rauf, Mark Fishel, Sandra Noubours, and Rico Sennrich. 2010. Evaluation of sentence alignment systems. Fifth MT Marathon. Le Mans, France.

Christopher Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA.

Robert C. Moore. 2002. Fast and accurate sentence alignment of bilingual corpora. In *AMTA '02: Proceedings of the 5th Conference of the Association for Machine Translation in the Americas on Machine Translation: From Research to Real Users*, pages 135–144, London, UK. Springer-Verlag.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computat. Linguist.*, 29(1):19–51.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL 2002*, pages 311–318, Morristown, NJ, USA.

Rico Sennrich and Martin Volk. 2010. MT-based sentence alignment for OCR-generated parallel texts. In *The Ninth Conference of the Association for Machine Translation in the Americas (AMTA 2010)*, Denver, Colorado.

Michel Simard, George F. Foster, and Pierre Isabelle. 1993. Using cognates to align sentences in bilingual corpora. In *CASCON '93: Proceedings of the 1993 conference of the Centre for Advanced Studies on Collaborative research*, pages 1071–1082. IBM Press.

A. Stolcke. 2002. SRILM – An Extensible Language Modeling Toolkit. In *Seventh International Conference on Spoken Language Processing*, pages 901–904, Denver, CO, USA.

Dániel Varga, László Németh, Péter Halácsy, András Kornai, Viktor Trón, and Viktor Nagy. 2005. Parallel corpora for medium density languages. In *Proceedings of the RANLP 2005*, pages 590–596.

Martin Volk, Noah Bubenhofer, Adrian Althaus, Maya Bangerter, Lenz Furrer, and Beni Ruef. 2010. Challenges in building a multilingual alpine heritage corpus. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

ISSN 1736-6305 Vol. 11
http://hdl.handle.net/10062/16955