

Bootstrapping Arabic-Italian SMT through Comparable Texts and Pivot Translation

Mauro Cettolo Nicola Bertoldi Marcello Federico

FBK - Fondazione Bruno Kessler
via Sommarive 18 - 38123 Povo, Trento, Italy
{cettolo, federico, bertoldi}@fbk.eu

Abstract

This paper describes efforts towards the development of an Arabic to Italian SMT system for the news domain. Since only very little parallel data are available for this language pair, we investigated both the exploitation of comparable corpora and pivot translation. Experimental evaluation was conducted on a new benchmark developed by extending two Arabic-to-English NIST evaluation sets. Preliminary results show potentials of both approaches with respect to performance achieved by a popular state-of-the-art Web-based translation service.

1 Introduction

Statistical machine translation (SMT) has shown to be a very competitive approach and almost language independent, at least in its basic setting. The concept that phrase-based SMT *only* needs one basic ingredient, parallel data, has propagated the belief that SMT systems can be *easily* ported across any language pair, given that sufficient parallel data are available. Such optimistic view has been also supported by prominent research projects, such as EuroMatrix,¹ and industrial efforts, mainly Google, which successfully developed SMT systems for a large number of translation directions. These initiatives however coped with SMT's data hunger in different ways. The EuroMatrix project took advantage of an existing and very popular multilingual parallel corpus (EuroParl) covering all official EU languages, while Google could rely on its impressive in-house Web

mining and computing infrastructure, which are doubtlessly unaffordable by most research labs.

Unfortunately, for most research labs, a major bottleneck that hinders the development of SMT systems for many language pairs is the lack of sufficient amounts of parallel data. In fact, parallel data are a scarce resource even for many socially and economically relevant language pairs,² such as Arabic and Italian.

In this work, we report on our efforts to set-up SMT baselines for news translation from Arabic to Italian, for which only a small amount of parallel texts was available. To compensate for the lack of data, we investigated and compared two alternatives: exploiting a fair amount of comparable data, automatically collected from the Web, and pivot translation through English. Results seem to reward the pivot translation and suggest further potential by combining the direct and pivot translation methods.

Experiments have been conducted on a new benchmark developed by the EUROMATRIXPLUS project,³ that extends two news data sets, employed in the 2009 Arabic-English NIST MT evaluation campaign. In particular, the Arabic side of the newswire evaluations sets was translated by experts both into French and into Italian.⁴ This benchmark was produced for the sake of fairly comparing direct versus pivot translation for two directions (Arabic-Italian/French) and two pivot languages (English and French), respectively. As of today, French has not been involved at all in our investigation: it will be considered in the near fu-

© 2011 European Association for Machine Translation.

¹www.euromatrix.net

Mikel L. Forcada, Heidi Depraetere, Vincent Vandeghinste (eds.)

Proceedings of the 15th Conference of the European Association for Machine Translation, p. 249–256
Leuven, Belgium, May 2011

²See for example: www.translated.net/en/languages-that-matter

³www.euromatrixplus.net

⁴These translations are publicly available through the project website.

ture.

The paper is organized in the following way: Section 2 reviews previous literature on the exploitation of comparable corpora for training SMT. Section 3 provides some more information on the NIST-based test sets mentioned above. Section 4 describes the methods we propose for mining parallel texts from multilingual Web sites and applied in experiments presented and discussed in Section 5. Some final remarks conclude the paper.

2 Related Work

Nowadays, several international news agencies deliver content through the Web in many languages. This represents a formidable opportunity to collect comparable documents, that is texts expressing the same content in different languages. Although these documents are not parallel, it often happens that some portions of them are mutual translations to some extent. In the last years, much effort has been devoted by the research community to the effective exploitation of such data for SMT.

The first problem to face is the alignment of multilingual documents reporting the same news. This problem has been rarely investigated systematically, as the usual pairing strategies aim to keep low the missing rate, that is to reward the recall. One of the most valuable methods is that presented in (Uszkoreit et al., 2010). There, all non-English documents are translated into English through a initial, even low-quality, translation system. Documents are then paired in two steps: the first generates a set of candidate pairs of documents sharing at least a certain number of rare features. This step is made linear in the number of input documents by setting a threshold defining such rare features. In the second step, a computationally more expensive and fine grained comparison is performed for deciding whether such document pairs are comparable or not.

Pairing multilingual documents is also the goal of (Steinberger et al., 2005), although the proposed method focuses more on the similarity of content rather than on the comparability of texts.

In Section 4.1 we propose three different document alignment methods and experimentally evaluate them on the basis of precision and recall. All methods share the initial translation of the document title, which is then used to detect comparable documents. In such respect, also our methods only rely on the textual content of the documents and

do not require any meta-data, much alike but more affordable than (Uszkoreit et al., 2010).

Once document pairs are available, the problem is the detection of actual parallel text inside them. Several approaches have recently been proposed in the literature to extract such parallel excerpts. Most of the techniques, if not all, share the stages of splitting documents into sentences and of pairing sentences across documents. Methods significantly differ in the successive filtering steps, and can be clustered into two main groups: procedures aiming at deciding if paired sentences are mutual translations or at detecting and extracting parallel sub-sentential fragments.

In (Munteanu and Marcu, 2005) words in the source documents are translated through a bilingual lexicon; all possible sentence pairs of the two documents are passed first through a word-overlap filter and then classified as parallel or not by a maximum entropy classifier trained on (a small amount of) parallel sentences.

In (Abdul-Rauf and Schwenk, 2009) a SMT system, trained on a small amount of parallel data, is used to directly translate the source side of the documents. Then, instead of the maximum entropy classifier, WER and TER scores are computed at the sentence level by comparing the translations with the target side; the amount of filtered pairs can be tuned by varying the acceptance threshold.

The approach in (Munteanu and Marcu, 2006) resembles (Munteanu and Marcu, 2005) up to the definition of the set of candidate sentence pairs. Instead of deciding whether the two sentences are mutual translations, now they search for parallel fragments using an approach inspired by signal processing. Using a set of parameters derived from Log-Likelihood-Ratio statistics, each word is annotated with values in $[-1, +1]$ indicating the likelihood that the word has some translations in the other sentence by performing a greedy alignment. This stream of values is then treated as a signal and passed through an averaging filter. Spans that have only positive signal values and are longer than a threshold (3 words) are considered more likely to have a translation on the other side. The same process is repeated on the other translation direction, and the resulting fragment pair is assumed to be parallel.

Quirk et al. (Quirk et al., 2007) try to overcome some of the limitations of the approach described in (Munteanu and Marcu, 2006) in the way paral-

set	#sentences	Arabic		English		French		Italian	
		$ W $	$ V $	$ W $	$ V $	$ W $	$ V $	$ W $	$ V $
eval08-NW	813	21.9k	7.8k	29.1k	4.9k	33.2k	4.9k	32.0k	5.7k
eval09-NW	586	17.5k	6.4k	23.1k	3.9k	26.7k	4.4k	25.1k	4.8k

Table 1: Statistics of the dev/test sets for the pivoting task. Texts are tokenized. For the English side, 4 manual translations (references) are available: average values are reported. $|W|$ stands for “running words”, $|V|$ for “vocabulary size”.

lel fragments are identified. In particular, they propose two generative models for generating noisy target sentences from the source sentences. One model is employed to align words in candidate sentence pairs; fragments are then extracted from alignments by applying simple heuristics. Another model tries to directly generate fragments; in the process, three generation options are competing: source-only fragment, target-only fragment, or joint source-target fragment. The rationale behind this model is that “the probability of generating source and target fragments jointly should be more likely than generating them independently if and only if they are parallel”. The latter model is definitely more complex than the former one, although their impact on SMT performance is similar.

In (Cettolo et al., 2010), a method is defined for extracting fragments which was new in some aspects and that is summarized in Section 4.2.

3 Evaluation Sets

Evaluation and comparison of MT systems are still open issues to which much attention and many efforts have been directed in the last years. One of the few certainties is the use of publicly available evaluation sets. Nowadays, several MT evaluation campaigns are held every year which release evaluation sets. Unfortunately, no evaluation set built on news texts has been made available so far for Arabic–Italian translation. To fill this gap, we decided to add two new target languages (Italian and French) to the development and test sets provided for the 2009 MT NIST evaluation campaign, for the Arabic to English newswire task. In order to ensure high quality, a professional translation company was in charge of translating the Arabic side of the two sets, in such a way to avoid any bias towards English. Just a single translation was requested for each sentence. Table 1 provides some statistics of the produced texts.

This benchmark, that is publicly released

through the EUROMATRIXPLUS project website, will be used to investigate direct and pivot translations from Arabic to Italian and French.

4 Comparable Corpora for SMT

The starting point for mining parallel text in any pair of languages is a collection of documents written in those two languages. The following subsections describe the main steps of our method to accomplish the task.

4.1 Document pairing

Once such a collection is at disposal, the first problem is to pair the documents which likely include parallel texts. In doing that, our twofold goal is to keep low both the number of false rejections, that is the missing of proper pairs, and that of false alarms, which unnecessarily activate the following processing steps.

Sometimes, the pairing is straightforward and reliable thanks to the presence of explicit links between documents reporting the same news in different languages. In less lucky cases, a widely adopted solution is to pair documents which have been published within a given window of days, for example just in the same day, or one day before or after.

In this work, we have investigated three more general methods, all assuming only the existence of a title in each news and the availability of a baseline MT system for the translation of titles from the language A into the language B. For each A title, the methods select the B documents that better match the A title. The publication date can be exploited if available, to limit the set of documents eligible to be paired, but its presence is not mandatory; if missing, the search is performed on the whole collection.

The three methods differ in the way the selection is performed.

In the first method, each A title is translated into B; then, its distance from the title of each B el-

igible document is computed; documents are finally paired if corresponding titles are closer than a given threshold θ . Any of the metrics for MT evaluation could be employed; for this specific task, we do not expect that the differences among them in terms of correlation with human judgments can significantly affect results. Hence, we decided in favor of Position-independent word Error Rate (PER), which is fast and works reasonably well.

The second method additionally translates the A title into B under the constraint defined by the B title; this is performed by means of the constrained phrase-based DP-search process described in (Cettolo et al., 2010). The algorithm provides the translation score and the coverage rate, that is the percentage of the source words actually translated. These two scores together with the PER from the first method are used to set up a Naive Bayes (NB) classifier with three features.

The third method is based on a standard IR technique: the set of B documents are first indexed and then retrieved with A titles translated into B. The document retrieval has been performed by means of Lucene (McCandless et al., 2010).

The three methods for document pairing sketched above have been tested on a set of 30k documents in Italian and 30k in English crawled from a news site, whose reference pairing was generated by exploiting explicit links. Concerning the θ and NB methods, the matching was attempted for documents whose publication dates differ for no more than 1 day. The resulting negative examples were 1.7 million, against 30k positive ones. For the IR method, the retrieval was performed only on documents published in the same day of the querying document (630k negative examples). Table 2 provides performance in terms of precision (P), recall (R) and F-score (F_1); in the first method, the threshold θ was chosen so as to maximize F_1 .

method	%P	%R	% F_1
θ	20.8	16.4	18.4
NB	26.8	25.3	26.0
IR	73.2	73.0	73.1

Table 2: Performance of the three investigated methods for pairing documents.

For the sake of comparison, the pairing of all the documents published in the same day yields a F_1 of 8% (93% of recall, 4% of precision); 3% (94%, 2%) within ± 1 day. Definitely, the inves-

tigated methods, and in particular that exploiting IR, result in a more accurate pairing than the blind one, based only on publication date. Nevertheless, it is not obvious whether maximizing the F-score is the optimal choice, given that each rejection avoid further processing, but the false ones prevent the detection of parallel texts.

4.2 Mining parallel fragments

This section summarizes the method we proposed in (Cettolo et al., 2010) for collecting parallel fragments of text from comparable documents, which is novel in two main aspects: (i) fragments are mined from document-sentence pairs rather than from sentence-sentence pairs; (ii) fragments are built on phrase- rather than word-level alignments. The approach comprises three steps: first, the text of the source documents is paired to each sentence of the target document; then, a partial phrase-based alignment between the paired texts is computed; finally, aligned phrases are iteratively merged into blocks on the basis of simple heuristics. Resulting aligned blocks represent the mined parallel fragments.

Translation experiments that assess the utility of the extracted fragments were conducted on the German-to-English task defined by the 2010 ACL SMT Workshop. Results reported in Table 3 show that by augmenting the baseline training data with parallel fragments mined from comparable in-domain news texts (NW), the BLEU score increases up to 5% relative (18.5 vs. 17.6). Adding the same amount of sentences from an out-of-domain parallel corpus (EuroParl, EP), BLEU remains below of more than 3% in relative terms (17.9 vs. 18.5); for reaching the same level of performance, more than four times of out-of-domain but parallel data is required.

training data			%BLEU
baseline W (source)	additional W type (source)		
	-	-	17.6
2.5M	0.5M	frag(NW)	18.5
	0.5M	sent(EP)	17.9
	2.0M	sent(EP)	18.3

Table 3: Performance comparison on German-to-English NewsTest2010 evaluation set by adding in-domain fragments vs. two different amounts of parallel out-of-domain sentences.

5 Arabic to Italian SMT Systems

In this section, the direct and pivot (via English) SMT systems for the translation of news from Arabic to Italian are presented and experimentally evaluated.

The automatic translators are built upon the open-source MT toolkit Moses (Koehn et al., 2007).⁵ The translation and the lexicalized re-ordering models have been trained on parallel data built as specified through the text. In all experiments, 6-gram LMs have been employed, smoothed with the improved Kneser-Ney technique (Chen and Goodman, 1999) and computed with the IRSTLM toolkit (Federico et al., 2008). The models are log-linearly interpolated, unless otherwise specified; the interpolation weights have been optimized on the development sets by means of the standard MERT procedure.

Development and test sets are those described in Section 3.

Section 5.1 provides details on each direct system involved in this experimental stage. Results on pivoting and their discussion are instead given in Section 5.2.

5.1 Direct systems

Table 4 collects the performance in terms of BLEU score on the eval09-NW set of the direct systems developed so far, namely the Arabic-to-Italian, the Arabic-to-English and the English-to-Italian systems. For the sake of comparison, performance of the Google Translate⁶ (suffix *ggle*) - as it was in January 2011 - are also reported, since it is widely considered a state-of-the-art MT system for any language pair. Of course, Google systems are not likely tuned for NIST benchmarks; anyway, the news domain is so vast that it is expected that general purpose MT systems well cover it.

A brief description of our direct systems follows:

Arabic-to-Italian (ArIt) The training parallel corpus used for the estimation of the translation and reordering models counts around 3M running words. It consists of sentences (1.4M words) and fragments (1.6M words) crawled from a number of Web sites: sentences from TED⁷ (Paul et al., 2010) are already parallel; documents from the other

⁵www.statmt.org/moses/

⁶www.google.com/language_tools

⁷www.ted.com

system id	lang. pair	%BLEU (4 refs)	%BLEU (1 ref)
ArIt	ar→it	-	13.1
ArIt-ggle		-	19.2
EnIt	en→it	-	21.0
EnIt-ggle		-	19.2
ArEn	ar→en	54.3	35.3
ArEn-ggle		55.5	33.5

Table 4: Performance on the eval09-NW set of the direct systems developed for the translation from Arabic into Italian via pivoting. Eval08-NW has been used for tuning.

sites have been aligned either through meta-data, or by means of the procedure described in Section 4.1. From the aligned documents, fragments have been extracted through the method sketched in Section 4.2. Preprocessing removed from parallel data too long sentences which would hurt the word alignment phase of the training process (entries marked as *clean* in tables).

In addition to the LM trained on the target side of the crawled text, a large LM has been built on more than 1G words from archives of Italian newspapers. Table 5 provides some statistics on texts employed for the estimation of SMT models.

type	W		trained models
	ar	it	
web parallel sent.	1.4M	1.4M	
web parallel frag.	1.8M	1.6M	
total		3.0M	LM
total clean	3.0M	2.8M	TM RM
web monol. sent.		1.06G	LM

Table 5: Training data for ArIt models.

English-to-Italian (EnIt) Unlike in the case of the Arabic-Italian pair, many Web sites publish news in English and Italian which are (almost) parallel. We have been able to collect so far sentences for a total of about 24M words per side. Document pairs have been split into sentences on strong punctuation and sentence alignment has been performed by means of Gargantua (Braune and Fraser, 2010). On the contrary, from sites where documents are at most comparable, fragments (2.8M words) have been mined via the procedures described in Sections 4.1 and 4.2. In addition,

Europarl⁸ and JRC-Acquis⁹ parallel corpora (70M words) have been employed. The LM of the `ArIt` system has been re-used. Table 6 provides some statistics on texts employed for the estimation of these specific SMT models.

type	W		trained models
	en	it	
web parallel sent.	24.2M	24.1M	
web parallel frag.	2.7M	2.8M	
total		27.0M	LM
total clean	23.3M	23.5M	TM RM
ep5+acquis clean	70.0M	70.0M	TM RM
web monol. sent.		1.06G	LM

Table 6: Training data for `EnIt` models.

Arabic-to-English (`ArEn`) The Arabic-to-English SMT system has been developed on the training data of the NIST 2009 MT evaluation campaign. Specific translation models were estimated on the 5 parallel corpora provided by the organizers and linearly interpolated to define a single model; on the target side, specific LMs were estimated as well; on the contrary, a single reordering model was built on such texts. Concerning the exploitation of monolingual data, again specific LMs were built on each of the 6 provided sets (English Gigaword 3rd Edition); also in this case, a single model has been defined by linearly interpolating the total of 11 LMs. Table 7 provides some statistics on texts actually employed for the estimation of SMT models.

corpus	W ar/en	trained models
GALE	0.9M	TM LM
Newswire	6.2M	TM LM
ISI	24.3M	TM LM
UN	115.2M	TM LM
total clean	142.2M	RM
Gigaword	3.6G	6 LMs

Table 7: Training data for `ArEn` models.

Looking again at Table 4, results clearly show that the amount of parallel texts available for training is critical: if enough, the development of state-

⁸www.statmt.org/europarl

⁹wt.jrc.it/lt/Acquis/

of-the-art SMT systems is possible (English-to-Italian and Arabic-to-English directions); otherwise (Arabic-to-Italian) the need arises for alternative approaches to achieve acceptable quality.

5.2 Pivoting

The pivoting technique employed here is the *composition*, also called *transfer* in (Wu and Wang, 2009), consisting in the translation of the source language into the pivot language, and of this into the target language. Table 8 shows the performance obtained by composing the proper direct systems presented above, together with that by composing the two corresponding Google Translate systems.

composition systems	lang. pair	%BLEU (1 ref)
<code>ArEn</code> \otimes <code>EnIt</code>	ar \rightarrow it	19.5
<code>ArEn-ggle</code> \otimes <code>EnIt-ggle</code>		18.2

Table 8: Performance on the eval09-NW set of the composition systems for the translation from Arabic into Italian.

The reported figures suggest two main cues:

- the pairing of a good system (`ArEn`, 35.3 %BLEU) and a quite reasonable one (`EnIt`, 21.0) yields acceptable performance on the Arabic-to-Italian task (19.5), equivalent to that of the state-of-the-art Google Translate (19.2 of Table 4), proving the effectiveness of the approach. In the next future, further efforts will be devoted in making stronger the `EnIt` SMT system;
- Google provides machine translation between any pair of 58 languages as of today and it is reasonable to believe that in many of them a pivot approach is used. However, pivoting through English with the two Google systems seems to achieve lower performance than the Arabic-to-Italian system of Google itself (18.2 vs. 19.2). This may have different explanations: e.g. a strong direct translation system, composition of several languages, or a different pivot approach. Anyway, this suggests us to focus future effort in making more effective our pivot chain from Arabic to Italian, for example by including more pivot languages. A good candidate is French, which is a language closer to Italian and for which much Arabic/French parallel data is available.

In our setup, the *triangulation* (Wu and Wang, 2009) approach to pivot translation, that is the multiplication of phrase tables, yielded a performance below the composition method (17.5 %BLEU). Nevertheless, it is worth noticing that by selecting sentence by sentence the best automatic translation (with respect to the TER score) among those generated by the direct system (13.1 %BLEU), the composition system (19.5) and the triangulation system (17.5), the final %BLEU is 20.7. The best sentence is picked up from the best system most of the times (58%) of course, but also the direct system contributes in a significant way (8% of the times). These figures show the high potential of system combination, which is a further direction of future research.

6 Final Remarks

The following comments hold for our experience in the development of the Arabic-to-Italian SMT system. They are not generalizable to any language pair, because there are languages for which either much larger amounts of data are available (e.g. Chinese-English) or even monolingual data are scarce (Pashto). Anyway, we think that excluding the extreme cases, our outcomes are valid for most language pairs.

Firstly, nowadays it is not easy to collect parallel data of quality and in quantity such that statistical models can be effectively trained. Then, the exploitation of pivot translation seems to be necessary. Nevertheless, consider that systems based on different approaches can provide good translations of distinct portions of text even when their overall quality differs a lot: this means that it can be profitable to develop many diverse systems - direct, pivot-based, etc. - and combine their outputs, independently from their quality.

Hence, our future work will focus on making the direct Arabic-to-Italian system stronger, on reinforcing the components of the pivot chain, by also adding more pivot languages to English, and on system combination.

7 Acknowledgements

This work was supported by the EuroMatrixPlus project (IST-231720), which is funded by the EC under the 7th Framework Programme for Research and Technological Development.

The authors also thank the three anonymous reviewers for their helpful comments.

References

- Abdul-Rauf, Sadaf and Holger Schwenk. 2009. On the use of comparable corpora to improve smt performance. In *Proc. of the 12th Conference of the European Chapter of the ACL (EACL)*, pages 16–23, Athens, Greece.
- Braune, Fabienne and Alexander Fraser. 2010. Improved unsupervised sentence alignment for symmetrical and asymmetrical parallel corpora. In *Coling 2010: Posters*, pages 81–89, Beijing, China.
- Cettolo, Mauro, Marcello Federico, and Nicola Bertoldi. 2010. Mining parallel fragments from comparable texts. In *Proc. of the International Workshop on Spoken Language Translation (IWSLT)*, pages 227–234, Paris, France.
- Chen, Stanley F. and Joshua Goodman. 1999. An empirical study of smoothing techniques for language modeling. *Computer Speech and Language*, 4(13):359–393.
- Federico, Marcello, Nicola Bertoldi, and Mauro Cettolo. 2008. IRSTLM: an open source toolkit for handling large scale language models. In *Proc. of Interspeech*, pages 1618–1621, Melbourne, Australia.
- Koehn, P., H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proc. 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.
- McCandless, Michael, Erik Hatcher, and Otis Gospodnetić. 2010. *Lucene in Action*. Manning Publications Co., 2nd edition.
- Munteanu, Dragos Stefan and Daniel Marcu. 2005. Improving machine translation performance by exploiting non-parallel corpora. *Computational Linguistics*, 31(4):477–504.
- Munteanu, D.S. and D. Marcu. 2006. Extracting Parallel Sub-Sentential Fragments from Non-Parallel Corpora. In *Proc. of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, pages 81–88, Sydney, Australia.
- Paul, Michael, Marcello Federico, and Sebastian Stücker. 2010. Overview of the IWSLT 2010 Evaluation Campaign. In *Proc. of the seventh International Workshop on Spoken Language Translation (IWSLT)*, pages 3–27, Paris, France.
- Quirk, Chris, Raghavendra Udupa, and Arul Menezes. 2007. Generative models of noisy translations with applications to parallel fragment extraction. In *Proc. of the Machine Translation Summit XI*, Copenhagen, Denmark.

- Steinberger, Ralf, Bruno Pouliquen, and Camelia Ignat. 2005. Navigating Multilingual News Collections using Automatically Extracted Information. *Journal of Computing and Information Technology*, 13(4):257–264.
- Uszkoreit, Jakob, Jay M. Ponte, Ashok C. Popat, and Moshe Dubiner. 2010. Large scale parallel document mining for machine translation. In *Proc. of the 23rd International Conference on Computational Linguistics (COLING)*, pages 1101–1109, Beijing, China.
- Wu, Hua and Haifeng Wang. 2009. Revisiting pivot language approach for machine translation. In *Proc. of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1, ACL '09*, pages 154–162, Stroudsburg, PA, USA.