Sub-sentential Paraphrasing by Contextual Pivot Translation

Aurélien Max
LIMSI-CNRS
Université Paris-Sud 11
Orsay, France
`aurelien.max@limsi.fr`

Abstract

The ability to generate or to recognize paraphrases is key to the vast majority of NLP applications. As correctly exploiting context during translation has been shown to be successful, using context information for paraphrasing could also lead to improved performance. In this article, we adopt the pivot approach based on parallel multilingual corpora proposed by (Bannard and Callison-Burch, 2005), which finds short paraphrases by finding appropriate pivot phrases in one or several auxiliary languages and back-translating these pivot phrases into the original language. We show how context can be exploited both when attempting to find pivot phrases, and when looking for the most appropriate paraphrase in the original sub-sentential "envelope". This framework allows the use of paraphrasing units ranging from words to large sub-sentential fragments for which context information from the sentence can be successfully exploited. We report experiments on a text revision task, and show that in these experiments our contextual sub-sentential paraphrasing system outperforms a strong baseline system.

1 Introduction

The ability to generate or to recognize paraphrases is key to the vast majority of NLP applications. Most current research efforts on paraphrase generation attempt to push the limits of their respective methods and resources without recourse to deep meaning interpretation, an admitedly long-term research objective. A step towards meaning-aware paraphrasing can be done by appropriate use of the context in which a paraphrasing occurrence occurs. At the lowest level, deciding automatically when a word can be substituted with a synonym is a complex issue (Connor and Roth, 2007). When attempting paraphrasing on a higher level, such as arbitrary phrases or full sentences (Barzilay and Lee, 2003; Pang et al., 2003; Quirk et al., 2004; Bannard and Callison-Burch, 2005; Zhao et al., 2008a), a first issue concerns the acquisition of elementary units, which in the general case do not exist in predefined dictionaries. Some paraphrasing strategy must then follow, which may consider the context of a substitution to guide the selection of appropriate units (Callison-Burch, 2008; Max, 2008). An important limitation to this family of works is the scarcity of corpora that can be used as reliable supervised training data. Indeed, strictly parallel sentence pairs, for instance, are not naturally produced in human activities.[1] As a consequence, works on paraphrasing have recourse to costly human evaluation procedures, and an objective of automatic evaluation metrics is to rely on as little gold standard data as possible (Callison-Burch et al., 2008).

A text revision task is an application of paraphrase generation where context may be used in an effective way. When a local change is made to a text, it occurs within a textual "envelope" within which a paraphrase should fit. In particular, if the original sentence was grammatical, the substituted sentence should remain grammatical and convey essentially the same meaning.[2] The manner in which such a context can be exploited depends of course on the type of automatic paraphrasing technique used. In this article, we adopt the pivot

---

[1] Recent works such as (Nelken and Yamangil, 2008) have proposed mining the revision histories of collaborative authoring resources like Wikipedia, offering interesting prospects in paraphrasing and rewriting studies.

[2] We posit here that the *revision* activity does not involve important semantic changes, as opposed to the *rewriting* activity. In future work, we will attempt to consider cases of paraphrasing involving meaning changes corresponding to textual entailment phenomena.

approach based on parallel multilingual corpora proposed by (Bannard and Callison-Burch, 2005), which finds short paraphrases by finding appropriate pivot phrases in one or several auxiliary languages and back-translating these pivot phrases into the original language. We show how context can be exploited both when attempting to find pivot phrases, and when looking for the most appropriate paraphrase in the original sub-sentential envelope. This framework allows the use of paraphrasing units ranging from words to large sub-sentential fragments for which context information from the sentence can be successfully exploited.

This article is organized as follows. In section 2, we briefly review related work in paraphrasing and context-aware Machine Translation. We describe the main characteristics of our approach to sub-sentential paraphrasing in section 3. We then describe an evaluation protocol for evaluating our proposal and report the results of a human evaluation in section 4. We finally conclude and present our future work in section 5.

2 Related work

Different sources have been considered for paraphrase acquisition techniques. (Pang et al., 2003), for example, apply syntactic fusion to multiple translations of individual sentences. (Barzilay and Lee, 2003; Dolan et al., 2004) acquire short paraphrases from comparable corpora, while (Bhagat and Ravichandran, 2008) considered the issue of acquiring short paraphrase patterns from huge amounts of comparable corpora. (Bannard and Callison-Burch, 2005) introduced a pivot approach to acquire short paraphrases from multilingual parallel corpora, a resource much more readily available than their monolingual counterpart. (Zhao et al., 2008b) acquire paraphrase patterns from bilingual corpora and report the various types obtained.[3] (Callison-Burch, 2008) improves the pivot paraphrase acquisition technique by using syntactic constraints at the level of constituents during phrase extraction. This works also uses syntactic constraints during phrase substitution, resulting in improvements in both grammat-

---

[3]The types of their paraphrase patterns are the following (numbers in parentheses indicate frequency in their database): phrase replacements (267); trivial changes (79); structural paraphrases (71); phrase reorderings (56); and addition of deletion of information that are claimed to not alter meaning (27).

icality and meaning preservation in a large-scale experiment on English. (Max, 2008) explored the use of syntactic dependency preservation during phrase substitution on French.

This family of works considered the acquisition of short paraphrases and their use in local paraphrasing of known units. Several works have tackled full sentence paraphrasing as a monolingual translation task relying on Statistical Machine Translation (SMT). For instance, (Quirk et al., 2004) used a phrase-based SMT decoder that uses local paraphrases acquired from comparable corpora to produce monotone sentential paraphrases. (Zhao et al., 2008a) acquired monolingual biphrases from various sources and used them with a phrase-based SMT decoder, and (Madnani et al., 2007) combined rules of their hierarchical decoders by pivot to obtain a monolingual grammar. These works were not motivated by the generation of high-quality paraphrases that could, for example, be reused in documents. The lack of structural information, the local nature of the paraphrasing performed and the fact that the context of the original sentences was not taken into account in the phrase-based approaches make it difficult to control meaning preservation during paraphrasing.

Context has been shown to play a crucial role in Machine Translation, where in particular proper Word Sense Disambiguation (WSD) is required in many cases. A variety of works have integrated context with some success into phrase-based and hierarchical decoders. For example, (Carpuat and Wu, 2007) disambiguate phrases using a state-of-the-art WSD classifier, and (Stroppa et al., 2007) use a global memory-based classifier to find appropriate phrase translations in context. Context is often defined as local linguistic features such as surrounding words and their part-of-speech, but some works have experimented with more syntactic features (e.g. (Gimpel and Smith, 2008; Max et al., 2008; Haque et al., 2009)).

Using an intermediate pivot language with bilingual translation in which a given language pair is low-resourced has led to improvements in translation performance (Wu and Wang, 2007; Bertoldi et al., 2008), but to our knowledge this approach has not been applied to full sentence paraphrasing. Several reasons may explain this, in particular the relative low quality of current MT approaches on full sentence translation, and the difficulties in controlling what is paraphrased and how.

3 Contextual pivot SMT for sub-sentential paraphrasing

Although many works have addressed the issue of local paraphrase acquisition, effective use of such paraphrases for paraphrase generation has only be achieved at the level of text units corresponding to short contiguous phrases. Recent works have proposed approaches to exploit context in order to correctly replace a text fragment with a paraphrase, but they are limited to known text units and therefore suffer from a scarcity of data.[4]

In this work, we address the case of sub-sentential paraphrase generation, an intermediate case between local paraphrasing using text units for which paraphrases are available and full sentence paraphrasing. Data sparcity is addressed by using a pivot translation mechanism, which can produce back-translations for text fragments for which paraphrases cannot be acquired beforehand by some paraphrase acquisition technique. Sub-sentential paraphrasing by pivot allows the exploitation of context during both source-to-pivot translation, where the source context is available, and during pivot-to-source back-translation, where the target context is known. The success of this approach is then directly dependent on the availability of high quality MT engines and on their ability to exploit these source and target contexts.

3.1 Paraphrasing by pivot translation

Whereas attempts at using two translation systems in pivot have met with some success for low-resourced language pairs, it is unlikely that current SMT systems can be successfully called in succession to obtain high-quality sentential paraphrases.[5] Several works have shown that monolingual biphrases obtained by multilingual pivots can be used by decoders, but although gains can for example be obtained by using sentential paraphrases as alternative reference corpora for optimizing SMT systems (Madnani et al., 2007), resulting paraphrases seem to be of too low quality for most other possible application contexts. In this work, we propose to use a pivot approach from a source language to a pivot language and back to the source language, but for sub-sentential fragments. In this way, the source context in which they occur can be exploited for both translating into the pivot language and for back-translating into the original language. This is illustrated on Figure 1.

Step (1) performs a N-best decoding (a single example is shown here) in which a segmentation of the source sentence is forced to ensure that a given fragment (*mettre en danger la richesse écologique* in the example) is translated independently of its surrounding context.[6] Only translations which respect this segmentation are kept, yielding a variety of pivot sentences. We are mostly interested in the pivot translation of our paraphrased fragment, but its prefix and suffix pivot context can be exploited by contextual SMT to guide pivot-to-source translation, although the lower quality of automatically generated sentences might not help as much as before.

Step (2) produces from each obtained N-best hypothesis a new N-best list of hypotheses, this time in the source language. The decoder is once more asked to use a given segmentation, and is further given imposed translations for the pivot prefix and suffix, as shown by the arrows going directly from the sentence at the top to the sentence at the bottom of Figure 1. Step (2) can be followed by a reranking procedure on the obtained N-best list of hypotheses, whose individual score can be obtained by combining the scores of the two translation hypotheses that led to it. As opposed to the pivot approach for phrases of (Bannard and Callison-Burch, 2005), it is not possible to sum over all possible pivots for a given pair ⟨original sentence, paraphrased sentence⟩, as the search space would make this computation impractical. We can instead look for the paraphrase that maximizes the product of the probabilities of the two translation steps according to the scores produced by the decoders used.

A further step can eliminate paraphrases by applying heuristics designed to define sought or undesirable properties for paraphrases, although this

---

[4]Current approaches based on paraphrase patterns are only a partial solution to this issue, as the variables used are limited to simple types.

[5]In particular, back-translation can introduce lexical errors due to incorrect word sense disambiguation and therefore severely hamper understanding, as illustrated by the infamous MT textbook example of the sentence *The spirit is willing but the flesh is weak* being translated into Russian and back-translated into English as *The vodka is good, but the meat is rotten.*

[6]It is in fact incorrect to say that translation of the various fragments would take place independently of each other, as various models such as a source context models or target language models will use information from surrounding fragments.
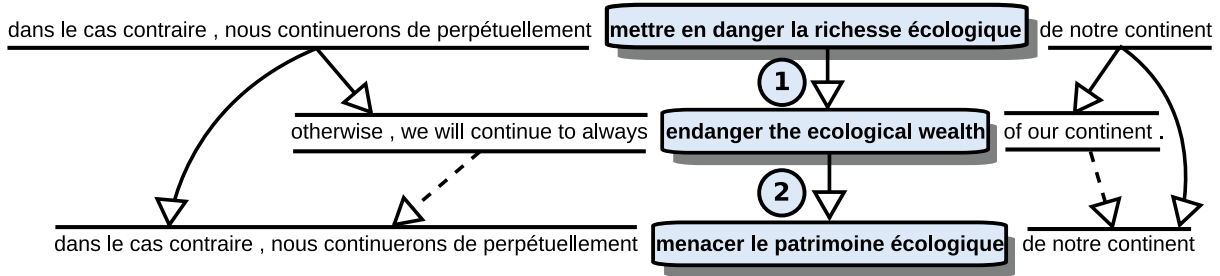
Figure 1: Example of sub-sentential paraphrasing by contextual pivot translation

could be directly integrated in the reranking step. For example, we may not be interested by identity paraphrases, or by paraphrases including or being included in the original fragment, or we may prefer paraphrases in which a given word has been replaced, etc.

3.2 Source context for pivot SMT

Using the context of a phrase is necessary to translate it correctly, most notably when several word senses corresponding to distinct translations are involved. The following examples show a case of a polysemous English word, which can be translated into three distinct French words and back-translated into various English fragments:

- *Follow the instructions outlined below to **save** that file. → sauvegarder ce fichier → write the file on disk*

- *Quitting smoking is a sure-fire way to **save** some money. → économiser de l'argent → have some money on your bank account*

- *Brown's gamble may **save** the banks but the economy cannot wait. → sauver les banques → salvage the banks*

Our approach for source context aware SMT, based on that of (Stroppa et al., 2007), is illustrated by the system architecture on Figure 2. A memory-based classification approach was chosen as it allows for efficient training with large example sets, can handle any number of output classes and produces results that can be directly used to estimate the required conditional probabilities. We add context-informed features to the log-linear framework of our SMT system based on the conditional probability of a target phrase e_i given a source phrase f_i and its context, $C(f_i)$:
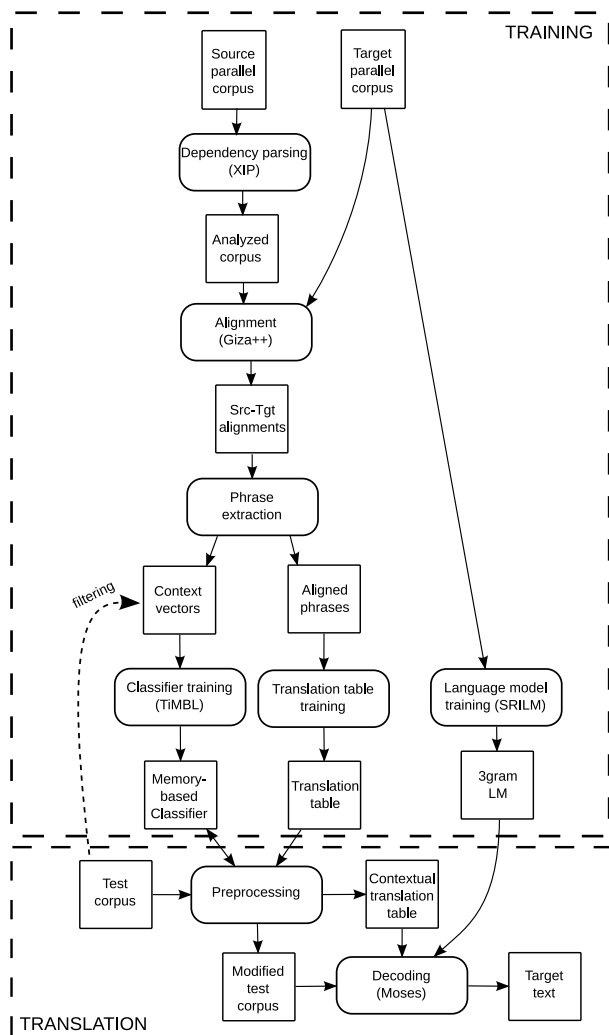
$$h_m(f_i, C(f_i), e_i) = \log P(e_i | f_i, C(f_i))$$

Figure 2: Architecture of our contextual phrase-based SMT system

21

Memory-based classification performs implicit smoothing, which addresses in part the problem of data sparcity, which worsen with the inclusion of context features and makes direct estimation of those probabilities problematic. Given a fixed-length vector, $\langle f_i, C(f_i) \rangle$, a set of weighted class labels corresponding to target phrases is returned by the classifier, which give access to $P(e_i | f_i, C(f_i))$ after normalization.

Because each source phrase potentially occurs in a unique context, they must be given a unique entry in the phrase table. To this end, we added a preprocessor component whose role is to dynamically build a modified source file containing unique tokens and to produce a modified translation table containing those tokens. Phrase extraction uses both phrase alignment results and linguistic analysis of the source corpus to produce standard biphrases and biphrases with contextual information. The latter are used to train the memory-based classifier. The source file undergoes the same linguistic analysis whose output is then aligned to unique tokens (e.g. president@45), and each possible phrase which is also present in the standard translation table is classified using its context information. The output is used to create a set of entries in the contextual translation tables, in which a new score corresponding to our context-based feature are added.

Most existing context-aware SMT approaches rely on context features from the immediate context of a source phrase. In this work, we initially restricted ourselves to a limited set of features: up to two lemmas to the left and to the right of a segment and their part-of-speech.[7]

3.3 Target context for pivot SMT

When decoding from the pivot hypothesis, we force our decoder to use provided sentence prefix and suffix corresponding to the "envelope" of the original fragment. Target context will thus be taken into account by the decoder.

Furthermore, based on the hypothesis that a paraphrase for an unmodified envelope should preserve the syntactic dependencies between the paraphrased fragment and its envelope (*inter-fragment dependencies*), we optionally add a "hard" reranking step where we filter the N-best list of hypotheses to keep only those which preserve these dependencies. Note however that for a dependency to be marked as preserved, we only need to find its label and its target word in the envelope (governor or dependent), as the word in the paraphrased fragment might have changed. This of course has practical implications on the nature of the paraphrases that can be produced.

In part due to various deficiencies of phrase alignments discussed in (Callison-Burch, 2008), we further apply heuristics to filter out some undesirable paraphrase candidates. Our current set of heuristics includes:

- no reordering should have taken place between the original source phrase and its context[8];

- considering the set of full word lemmas for the original fragment and the paraphrased fragment, at least one lemma should not belong to both sets[9];

- neither the original fragment nor its paraphrase must be included into the other (only taking full words into account).

4 Experiments

We have conducted experiments motivated by a text revision task that we report in this section by describing our baseline and context-aware subsentential paraphrasing systems and the results of a small-scale manual evaluation.

4.1 Data and systems

We built two-way French-English SMT systems using 188,115 lines of the Europarl corpus (Koehn, 2005) of parliamentary debates with moses (Koehn et al., 2007) [10]. Our corpus was analyzed by the XIP robust parser (Aït-Mokhtar et al., 2002) and its output tokenization was used. We built standard systems, as well as a contextual system for French→English as described in section 3.2 using an additional contextual score ob-

---

[7]We will integrate richer syntactic context as in (Gimpel and Smith, 2008; Max et al., 2008) in our short-term future work, as we expect it to be particularly useful for our paraphrasing task.

[8]Reordering is allowed in the paraphrased fragment.

[9]As a consequence, minimal paraphrases may differ by only one full word. This can however be used advantageously when the sought type of paraphrasing aims at "normalizing" a text fragment and when the most appropriate rewording is very similar to an original text fragment.

[10]We used revision 2234 available on the moses SVN website: `http://mosesdecoder.sourceforge.net/svn.php`. In particular, it allows the use of XML annotations to guide the translation of particular fragments.

Baseline fr→en	30.56
Contextual fr→en	31.17
Baseline en→fr	32.10

Table 1: BLEU scores for the translation systems used by our paraphrasing system

tained through memory-based classification performed with the TiMBL package (Daelemans et al., 2007). Standard MERT was used to optimize model weights. BLEU scores for the three systems are reported on Table 1. The contextual system obtains a slightly higher score than the baseline system, which can participate to some extent to a better exploitation of context for paraphrasing.[11]

Two paraphrasing systems we built: S_{bas} is a baseline system which uses standard phrase tables and post-filtering heuritics, but does not include reranking based on syntactic dependencies. S_{cont} is a contextual system which uses the contextual French→English translation system, reranking based on syntactic dependencies and post-filtering heuristics.

We used 1000-best lists of hypotheses for the source-to-pivot translation, and restricted ourselves to much smaller 10-best lists for pivot-to-source translation (integrating early more constraints directly into decoding could help in obtaining better and smaller *N*-best lists).[12]

4.2   Evaluation protocol

A native speaker was asked to study a held-out test file of Europarl data in French and to identify at most one fragment per sentence that would be a good candidate for revision and for which the annotator could think of reasonable paraphrases that did not involve changes to the envelope. Candidate fragments were accepted if they were not found in the French→English translation table. This step resulted in a corpus of 151 sentences with as many test fragments, with sizes ranging from 2 to 12 words, an average size of 5.38 words and a median size of 4 words.

Two native speakers, including the previous annotator, were asked to evaluate all paraphrased sentences on grammaticality and meaning. Contrary to previous works, we decided to use a

---

[11]The unexpected worse performance of the fr→en system may be explained by issues related to tokenization after analysis by the parser.

[12]In our future work, we intend to investigate the possible use of lattices rather than *N*-best lists.

smaller evaluation scale with only 3 values, as using more values tend to result in low inter-annotator agreement:

- 2: indicates that the paraphrase is perfect or almost perfect;

- 1: indicates that the paraphrase could become grammatical with one minor change, or that its meaning is almost clear;

- 0: indicates that more than one minor change is required to make the paraphrase grammatical or understandable.

4.3   Results and discussion

We ran both systems and took their one-best hypothesis for evaluation. Table 2 shows the results of a contrastive evaluation of the results obtained. For the 143 sentences for which paraphrases could be produced, we obtained 72 results common to both systems, and 71 which were specific to each system. The fact that for half of the cases the two systems produced the same paraphrases reveals that either context did not play an important role in these cases, and/or that the search space was rather limited due to the presence of rare words in the original fragment. Systems S_{cont} and S_{bas} are compared based on the number of cases were one was found to be better or worse than the other for the 71 cases where they proposed different paraphrases, either by the two judges (denoted by the $<$ and $>$ signs) or by one of the two judges while the other found the two systems to be of comparable performance (denoted by the $\leq$ and $\geq$ signs). As can be seen from the table, there is a clear preference for our S_{cont} system, with a 31:37 ratio of cases where it is preferred for grammar, and 33:49 for meaning.

Table 3 shows absolute results for the same run of evaluation. First, it can be noted that both systems perform at a reasonable level, both for short and long text fragments. Several reasons may explain this: first, sentences to paraphrase are from the same domain as the training corpora for our SMT systems, which is a positive bias towards the paraphrasing systems. Also, the post-filtering heuristics and the fact that both systems could benefit from the knowledge of the target envelope during pivot-to-source back-translation certainly helped in filtering out incorrect paraphrases. These results confirm the trend observed on contrastive results that our S_{cont} system is the best

	$S_{cont} < S_{bas}$	$S_{cont} \le S_{bas}$	$S_{cont} \ge S_{bas}$	$S_{cont} > S_{bas}$	?	Total
Grammar	3	3	10	21	34	71
Meaning	3	13	13	20	22	71

Table 2: Contrastive results. The notation $S_{cont} < S_{bas}$ stands for cases in which S_{cont} is found to be worse than S_{bas} by both judges; $S_{cont} \le S_{bas}$ for cases where S_{cont} was found to be worse by one judge while the other found the two systems equivalent; similarly for other cases. '?' stands for cases where judges disagreed.

	count	Grammar			Meaning			System		
		-	+	?	-	+	?	-	+	?
S_{bas} and S_{cont}	72	0	69	3	1	67	4	0	66	6
S_{bas} only	71	13	46	12	18	41	12	9	39	23
S_{cont} only	71	5	63	3	8	56	7	4	55	12
S_{bas}: $2 \le$ size ≤ 5	81	6	69	6	10	63	8	4	61	16
S_{cont}: $2 \le$ size ≤ 5	81	2	78	1	6	72	3	2	71	8
S_{bas}: $6 \le$ size ≤ 12	62	7	46	9	9	45	8	5	44	13
S_{cont}: $6 \le$ size ≤ 12	62	4	54	4	3	51	8	2	50	10

Table 3: Absolute results for manual evaluation. '+' indicates that both judges agree on a positive judgement (score 1 or 2), '-' that both judges agree on a negative judgment (score 0), and '?' that judges disagreed. 'System' judgments include judgments for both Grammar and Meaning.

performer for that task and that test set. It is however noteworthy that results were significantly better when they were produced by both systems, which may correspond to the easiest cases with respect to the training data and/or the task but also suggests the application of consensus techniques as done in MT system output combination.

Table 4 shows paraphrasing examples produced by S_{cont}. As can be noted from positive examples (a-c), the obtained paraphrases are mostly of the same syntactic types as the original phrases, which may be due to the proximity between the main language and the pivot language, as well as to the constraint on syntactic dependency preservation. Example (a) shows a case of what may be seen as some sort of normalization, as the concept of "*confidence of people*" (w.r.t. the English pivot language) may be more frequently expressed as *la confiance des citoyens (citizens)* than as *la confiance des gens (people)* in the reference corpus. Example (b), although showing a correct paraphrasing, contains an agreement error which is a result of the use of the gender neutral English pivot and the fact that the language model used by the pivot-to-source SMT system was not able to choose the correct agreement. Example (c) illustrates a case of correct paraphrasing involving reordering strongly influenced by the reordering re-

quired by the pivot language. The incorrect paraphrase of example (d) mainly results from the inability of the source-to-pivot SMT system to correctly translate the selected fragment; in particular, the syntactic structure was not correctly translated, and the noun *palier (stage)* was incorrectly translated as the verb *heal* and back-translated as the verb *traiter (heal, cure)*. Lastly, example (e) contains an error in word sense disambiguation between the pivot noun *act* and the noun *loi (law)*[13], as well as the incorrect deletion of the adverb *très fermement (firmly)* during source-to-pivot translation.

Several conclusions can be drawn from these results and observations. First, it is not surprising that the performance of the SMT systems used has an important impact on the results. This can be mitigated in several ways: by attempting paraphrasing on in-domain sentences; by using an appropriate pivot language with respect to the nature of the text fragment to paraphrase; by using one or several pivot languages (as proposed by (Bannard and Callison-Burch, 2005) for phrase paraphrasing) and consensus on the obtained paraphrases.

---

[13]This example might call for better lexical checking between original and paraphrased sentences, as well as exploiting context-aware SMT on the lower quality input of pivot-to-source translation.

(a)	En tant que parti de gauche, nous avons dû, avec beaucoup de peine, nous rendre compte que les institutions ne sont pas des jeux de construction montables, transformables et démontables à souhait, mais qu'elles ont leur propre histoire et **doivent bénéficier de la confiance des gens** qui les soutiennent. As the left, we have had, with a great deal of trouble, we see that the institutions are not games montables construction, transformables démontables and to wish, but they have their own history and **must enjoy the confidence of people who support them.** → **doivent avoir la confiance des citoyens**
(b)	Monsieur le président, **je suis inquiète au sujet de** l'attitude qui risque de se développer au sein de l'UE concernant la liberté des échanges. Mr President, **I am concerned about** the attitude which might develop within the EU on free trade. → **je suis préoccupé par**
(c)	Ces accords constituent **un cadre contractuel entièrement nouveau** pour les pays de la région. These agreements constitute **an entirely new contractual framework** for the countries of the region. → **un tout nouveau cadre contractuel**
(d)	Aujourd'hui, le durcissement parallèle des indépendantistes albanais et des autorités serbes **fait franchir à la crise un nouveau palier très inquiétant** dans la montée des tensions. Today, the inflexibility parallel with the Albanian independent and the Serbian authorities **to overcome the crisis is a new heal very worrying** in the rise of tension. → (*) **de surmonter la crise est une nouvelle traiter très préoccupant**
(e)	La commission **condamne très fermement cet acte** et prend note de la décision de constituer un comité spécial au sein de la fiscalia general de la nación afin d'enquêter sur cet assassinat. The Commission **condemn this act** and takes note of the decision to set up a special committee fiscalia within the general de la nacin in order to investigate this murder. → (*) **condamne cette loi**

Table 4: Examples of sub-sentential paraphrasings produced by our S_{cont} system.

Another remark is that our systems could be improved as regards their ability to exploit source context.[14]

5 Conclusion and future work

In this article, we have presented an approach to obtain sub-sentential paraphrases by using pivot SMT systems. Although our results showed that we were able to build a strong baseline on our test set, they also showed that integrating context both when translating from source-to-pivot and when back-translating from pivot-to-source can lead to improved performance. Our approach has the distinctive feature that it targets text fragments that can be larger than phrases traditionally captured by statistical techniques, while not targeting full sentences for which it would be harder to exploit context successfully. More generally, it addresses the case of the paraphrasing of text units for which no paraphrases are directly available.

We have identified several issues in our experiments that will constitute our future work. We intend to experiment with several pivot languages and to make them compete to obtain the *N*-best lists, as done in some approaches to multisource translation (Och and Ney, 2001) and/or to use a consensus technique to select the best paraphrase.

As regards our context-aware SMT systems, we plan to exploit more complex syntactic knowledge and to learn correspondances for inter-fragment dependencies so as to make our rescoring based on syntactic dependencies more flexible. We are currently working on extracting revision instances from Wikipedia's revision history, which will provide us with a corpus of genuine revision occurrences as well as with an out-domain test corpus with reference paraphrases. Lastly, we want to investigate the use of our approach for two types of applications: text normalization, in which a text is revised to select approved phraseology and terminology, through the use of a carefully chosen training corpus for our pivot-to-source SMT system ; and interactive translation output revision for cases with or without a source text for professional translators or monolingual users.

Acknowledgments

References

Salah Aït-Mokhtar, Jean-Pierre Chanod, and Claude Roux. 2002. Robustness beyond shallowness: incremental deep parsing. *Natural Language Engineering*, 8(3):121–144.

Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In *Proceedings of ACL*, Ann Arbor, USA.

---

[14]It should be noted, however, that the reported experiments used relatively small amounts of training data as in most comparable works on context-aware Machine Translation.

Regina Barzilay and Lillian Lee. 2003. Learning to paraphrase: an unsupervised approach using multiple-sequence alignment. In *Proceedings of NAACL/HLT*, Edmonton, Canada.

Nicola Bertoldi, Madalina Barbaiani, Marcello Federico, and Roldano Cattoni. 2008. Phrase-based statistical machine translation with pivot languages. In *Proceeding of IWSLT*, pages 143–149, Hawaii, USA.

Rahul Bhagat and Deepak Ravichandran. 2008. Large scale acquisition of paraphrases for learning surface patterns. In *Proceedings of ACL: HLT*, Columbus, USA.

Chris Callison-Burch, Trevor Cohn, and Mirella Lapata. 2008. Parametric: An automatic evaluation metric for paraphrasing. In *Proceedings of COLING*, Manchester, UK.

Chris Callison-Burch. 2008. Syntactic constraints on paraphrases extracted from parallel corpora. In *Proceedings of EMNLP*, Hawai, USA.

Marine Carpuat and Dekai Wu. 2007. Context-dependent phrasal translation lexicons for statistical machine translation. In *Proceedings of Machine Translation Summit XI*, Copenhagen, Denmark.

Michael Connor and Dan Roth. 2007. Context sensitive paraphrasing with a single unsupervised classifier. In *Proceedings of ECML*, Warsaw, Poland.

Walter Daelemans, Jakub Zavrel, Ko van der Sloot, and Antal van den Bosch. 2007. TiMBL: Tilburg Memory Based Learner, version 6.1, Reference Guide. Technical report, ILK 07-xx. Available from http://ilk.uvt.nl/downloads/pub/papers/ilk0703.pdf.

Bill Dolan, Chris Quirk, and Chris Brockett. 2004. Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources. In *Proceedings of Coling 2004*, pages 350–356, Geneva, Switzerland.

Kevin Gimpel and Noah A. Smith. 2008. Rich source-side context for statistical machine translation. In *Proceedings of WMT at ACL*, Columbus, USA.

Rejwanul Haque, Sudip Kumar Naskar, Yanjun Ma, and Andy Way. 2009. Using supertags as source language context in smt. In *Proceedings of EAMT*, Barcelona, Spain.

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

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of MT Summit*, Phuket, Thailand.

Nitin Madnani, Necip Fazil Ayan, Philip Resnik, and Bonnie J. Dorr. 2007. Paraphrases for parameter tuning in statistical machine translation. In *Proceedings of Workshop on Machine Translation at ACL*, Prague, Czech Republic.

Aurélien Max, Rafik Makhloufi, and Philippe Langlais. 2008. Explorations in using grammatical dependencies for contextual phrase translation disambiguation. In *Proceedings of EAMT*, Hamburg, Germany.

Aurélien Max. 2008. Local rephrasing suggestions for supporting the work of writers. In *Proceedings of GoTAL*, Gothenburg, Sweden.

Rani Nelken and Elif Yamangil. 2008. Mining wikipedia's article revision history for training computational linguistics algorithms. In *Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy*, Chicago, USA.

Franz Josef Och and Hermann Ney. 2001. Statistical multi-source translation. In *Proceedings of MT Summit*, Santiago de Compostela, Spain.

Bo Pang, Kevin Knight, and Daniel Marcu. 2003. Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences. In *Proceedings of NAACL/HLT*, Edmonton, Canada.

Chris Quirk, Chris Brockett, and William B. Dolan. 2004. Monolingual machine translation for paraphrase generation. In *Proceedings of EMNLP*, Barcelona, Spain.

Nicolas Stroppa, Antal van den Bosch, and Andy Way. 2007. Exploiting source similarity for smt using context-informed features. In *Proceedings of TMI*, Skvde, Sweden.

Hua Wu and Haifeng Wang. 2007. Pivot language approach for phrase-based statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic.

Shiqi Zhao, Cheng Niu, Ming Zhou, and Sheng Li. 2008a. Combining multiple resources to improve smt-based paraphrasing model. In *Proceedings of ACL-HLT*, Columbus, USA.

Shiqi Zhao, Haifeng Wang, Ting Liu, and Sheng Li. 2008b. Pivot approach for extracting paraphrase patterns from bilingual corpora. In *Proceedings of ACL-HLT*, Columbus, USA.