A Hybrid Model for Autonomous Danish-Arabic Statistical Machine Translation

Mossab Al-Hunaity

Abstract: We present a simple and efficient method for enhancing the Danish-Arabic (DA-AR) statistical machine translation system. The model mainly is composed of two major parts, information retrieval unit and SMT system. We train our baseline with small DA-AR corpora. We use the Arabic translation output as a query to Lemur information retrieval tool to search for a similar matching sentence in a very larger Arabic corpus. We use Translation Error Rate (TER) filter to select the best output of the IR system. We evaluate our approach and prove that it enhances the quality of translation. We extend our experiments to measure the effect of adding more language resources to our baseline. We mine available DA-EN and EN-AR resources to produce parallel DA-AR sentences. We use the new resources in training our baseline. We evaluate the quality of the extracted data by showing that it significantly improves the performance of our baseline performance.

Keywords: (DA-AR), (TER), Danish-Arabic, DA-EN and EN AR, baseline performance

I. INTRODUCTION

Developing a statistical machine translation (SMT) system for languages with limited common bilingual resources like the case with Arabic (AR) and Danish (DA) is a challenging task. To produce good results SMT system needs huge amounts of bilingual data to produce reasonable translations. The challenging question is how our SMT system can produce a reasonable translation without having enough bilingual training data. To answer this question we started from the observation that, for a domain specific text, like the computer printer manuals, usually sentences are repeated through many different printer manuals. Sentences will have similar structure and it's very common to find similar sentences between many printer manuals with the same meaning and structure, like for example: "turn on your printer" or "insert installation desk in computer drive", etc. Now if we have an SMT system that is trained on small amounts of language resources then the output translated sentence would have a weak syntactic and grammatical structure, but still normally it has the meaning. So we use this fact to correct our SMT translation. For example, if we were to translate between two languages with limited resources like Swahili and English where Swahili is the source language and English is the target, and let's assume that we are trying to translate a computer printer manual. It is very likely to produce unreasonable translation so we might have an output translation like "printer insert disk", but if we search our English manual we will find that the most suitable sentence that is composed of similar numbers of Swahili words and have "printer insert disk" would be "Insert printer installation disk into drive".

Manuscript Received on June 2014.

Mossab AL Hunaity, Department of Computer Information Systems, Ajman University, Ajma, United Arab Emirates. E-mail: <u>mossab99@gmail.com</u> Mobile: +962777496676

So we search for the possible and correct translation in the target language corpus to find the correct translation. We argue that if you have a very large domain specific corpus for a target language, then it is very likely that you will have a similar sentence that matches the meaning of your baseline translation. We will apply the same idea on our DA-AR case. So we build a baseline that is trained on a small bilingual corpus and a DA-AR dictionary where Danish is the source language and Arabic is the target language. We utilize the Lemur information retrieval tool to search for all possible similar sentences in a very large Arabic corpus that has that is domain specific to the baseline corpus. We managed to find correct translations similar to our baseline translated output. This approach enabled us to overpass the obstacle of having limited resources between Danish and Arabic. Although the system may not produce literal translation in many cases, it still produces a comparable translation that is syntactically and grammatically correct and still conveys the same meaning as the original sentence. And we think this is how human translation tends to be, it is more "meaning oriented" rather than "literal oriented" process.

We also try to enhance our baseline performance by mining all possible parallel DA-AR text from DA-EN, and EN-AR resources that are freely and widely available. We retrain our baseline with the extracted data. We report performance improvements on our system BLEU scores. In the next section we describe related work. Section 3 presents our system description. In section 4 we describe our data resources. We present our system architecture in section 5. In section 6 we present and explain the results of our experiments. Finally we discuss our conclusions and future work in section 7.

II. RELATED WORK

The idea of using information retrieval techniques to enhance SMT systems was introduced by many researchers, for example, Oard (1997), Och et.al (2002,2003). Eck et al. (2005) they proposed a new technique to select matched sentences based on n-gram coverage, n-grams were used to figure out how important the sentence was. The selected the frequency of the n-gram appearance is regarded as an indicator for selection, unseen n-grams were favored for selection. TF-IDF weighting scheme was tested in this method but no improvements was reported over n-grams, the idea was to decrease the amount of training data so that to make it possible to be deployed for small devices like PDA. Hildebrand et al. (2005) used information retrieval method for translation matching, they start with an adaptable translation model and they select similar sentences from a test set from in-domain and out-of domain training data.

> Published By: Blue Eyes Intelligence Engineering & Sciences Publication Pvt. Ltd.



A Hybrid Model for Autonomous Danish-Arabic Statistical Machine Translation

We use the same approach almost with a concentration on "in-domain" training data to detect a candidate translation in the in-domain text. Zhao et al. (2004) and Eck et al.(2004) they developed many experiments to use information retrieval as an adaptation model for supporting their SMT systems. Zhang et al (2006) and Mauser et al. (2006) use a new modified language model for their SMT, where one translation is used for re-ranking the translation output, the language model is built for the target language for the SMT, n-best translation candidate is generated after the first pass for the baseline SMT. In our work we try to boost also system performance by extraditing parallel Danish-Arabic sentences from available resources. Many researches inspected the problem of enhancing the performance of SMT systems by discovering parallel sentences from different linguistic resources. Resnik and Smith (2003) propose their STRAND web-mining based system for mining parallel sentences. The system manages to find large number of similar documents. Utiyama and Isahara, (2003) also developed another interesting technique for finding parallel sentences from comparable corpus between Japanese and English using dynamic programming and cross language information retrieval methods. Their approach was to identify similar article pairs and then treat these pairs as parallel text, then they tried to find similar sentences with similarity score and document pair measures , they would declare a match based on the least-cost alignment over the document pair measure. We use a similar approach but for finding a similar translation based to our baseline output. Yang and Lee (2003) they use dynamic programming to find parallel sentences in title pairs, they calculate confidence score to find a match between sentences which is based on longest common subsequence, In our work we use a similar approach for finding a matching sentence pair. Fung and Cheung (2004) used the same IR approach, they deployed the cosine similarity to match documents, they worked on noisy comparable corpora, their method will generate many candidate parallel selections where the best match is selected based on a threshold of cosine similarity scores, the use the extracted sentences would serve to build a dictionary to enhance their SMT system. Munteanu and Marcu (2005) uses bilingual dictionary to translate some of the words of source sentences. These translations are then used to query another system to find a matching translation using IR techniques. Candidate sentences will be selected based on words overlap. The maximum entropy classifier is used to select the best matching parallel sentences. Increasing the size of the bilingual dictionary and the Bootstrapping method is used to produce better results. Our technique is similar to that of Munteanu and Marcu (2005) but we don't use a dictionary for sentence discovery, instead we use our baseline output translation as a reference for our sentence matching query. Abdul-Rauf, Schwenk (2009) suggested a similar approach to Munteanu and Marcu, (2005), They extract parallel sentences from comparable corpora using IR techniques and then use that as an input for their SMT baseline, the translation is used again with another IR tool to find parallel sentences. Our work is differs from their approach in that we don't target parallel sentences rather than we search for similar comparable sentences as a target translation.

III. BASELINE DESCRIPTION

Our system is based on the Moses SMT toolkit (Koehn et al., 2007), we intend to translate from Danish to Arabic. The system is constructed as follows. First, we use Giza++ to perform word alignments in both directions. Second we extract phrases and lexical re-orderings using the default settings of the Moses SMT toolkit. The 4-gram back-off target LM is trained on the 3.0 Giga words Arabic monolingual Gigaword corpus. The Danish text we intend to translate to Arabic shares the same nature as the Arabic monolingual corpus therefore, it is likely that the target Arabic language model includes at least some of the translations of the Danish text, section 4 explains experiment data in more details. We argue that this is a key factor to obtain good quality translations. The translation model was trained on the Arabic Novels translated to Danish corpus (1.3 M words) and a DA-AR dictionary of about 500k entries. In a different version of this system, more language resources like Europarl¹ (47M words), Acquis² corpus (31M words), United Nations ³ corpus, Meedan ⁴ translation memory and LDC⁵ (catalog no. LDC2004T17) were added later to our experiments to boost baseline performance. The system interacts with Lemur information retrieval system, see section 5, and use the output of the text mining engine that we develop to increase our DA-AR parallel resources.

IV. DATA

In our experiment we use two sources of data; parallel DA-AR resources and comparable DA-EN, EN-AR language resources. For our baseline we need a common bilingual DA-AR data source. Unfortunately parallel DA-AR resources are not very common, but we managed to develop a 1.3 M parallel DA-AR corpus. It is based on Arabic novels that were translated for Danish. We divide this corpus into two sets; training set (1M) and testing set (300 K). These novels were written in Modern Standard Arabic. Data were processed for typing mistakes and were tokenized before being used. We also prepared a 1M word monolingual Danish corpus that is similar in text to the training data for testing purposes as well.

For our language model, we collected a 3.0 Giga word monolingual Arabic corpus that shares the same text (Novels) with the training data. Data was processed for typing mistakes and text was tokenized the same way we did with training data. Table 1 explains our experiments data sources. It is important to notice that the Arabic language model includes somehow the translation of the Danish sample test data. We argue that this is a key factor in our experiment to have a quality of translation.

The baseline will produce a translation for the input Danish sentence. This Arabic translation might contain problems like grammatical or syntactical errors or word ordering problems (Verb, SUB, OBJ). We solve this shortcoming by searching for a similar sentence in the large Arabic corpus. Later we output the most relevant Arabic sentence to the baseline translated Danish sentence to be a possible translation for the system instead of the baseline original translation. The benefit of this approach is that we will utilize the Arabic LM corpus to search for a suitable translation.

> Published By: Blue Eyes Intelligence Engineering & Sciences Publication Pvt. Ltd.



We are aware of the fact that novels as a text are not the best source of data you can get for your SMT system. A better source might be as the one presented by our introduction example, a domain specific text like hardware manuals, climate and whether change reports, etc. but in our case where only limited resources are available between Danish and Arabic we need to invest in these available resources. We selected our test data carefully to make sure that the content of the test data is contained in the large Arabic monolingual corpus so that we can test our theory. A major goal we are trying to achieve here is how we can extract knowledge (translation) out of noisy environment. We discuss our approach in details in section 5.

Another factor we try to inspect in our experiment is the additionalthe additional out of domain mined data in our baseline performance, for this purpose we will make use of the available comparable language resources for the two pairs of languages DA-EN and EN-AR. For the DA-EN group we will be using Acquis and Europarl parallel corpora. Acquis is collection of legislative texts from the European Union (EU) member states parliament meetings, while the Europarl parallel corpus is extracted from the proceedings of the European Parliament. Both Acquis and Europarl data domain is formal and covers the legal issues and they are free to use. For the Arabic-English pair we selected three major resources, the United Nations (UN) multilingual corpus, Meedan translation memory and LDC (catalog no. LDC2004T17). Both UN and Meedan are free to use and download. The domain of this group is mainly news, table 1 explains more details about our data resources. 1: http://www.statmt.org/europarl/

- 1: http://www.statmt.org/europarl/
- 2: Acquis http://langtech.jrc.it/JRC-Acquis.html
- 3: UN Corpus http://www.uncorpora.org/
- 4:Meedan http://github.com/anastaw/Meedan-Memory
- 5: LDC http://www.ldc.upenn.edu/

Both groups (DA-EN, EN-AR) inter cross partially in domain, so we can classify the two groups as a partially comparable resources which will be used to extract parallel Danish-Arabic sentences. Later this data will be used to boost our baseline performance. Section 5.3 will provide more details about this process details.

Fable 1: Data	Resources
---------------	-----------

Name	Direction	Domain	Size (words)
Training Corpus	Danish- Arabic	Arabic Novels	1.3 M
LM Corpus	Arabic	Arabic Novels	3.0 G
Europarl	Danish- English	Legal	47 M
Acquis	Danish- English	Legal	31 M
UN multilingual corpus	Arabic- English	Legal issues / News	3.2 M
Meedan	Arabic- English	News	0.5 M
LDC2004T17 Arabic- English		News	0.5 M

V. SYSTEM ARCHITECTURE

The general architecture of our system is shown in figure 1. Starting from a small parallel AR-DA corpus and an AR-DA dictionary we build a basic baseline see Section 3 for details.. To ensure a correct and accurate translation we use system output translation which may have syntactical or grammatical problems to search a very large corpus for similar sentences. We use Lemur information retrieval system to extract the most similar sentence to our baseline translation output.



Figure 1: System Major Architecture

We check the quality of our selected sentence using simple metrics like Word Error Rate (WER) and Translation Error Rate (TER) which will help us filter out good sentence matching pairs. If we find a good matching sentence then we replace the original baseline sentence translation with the new found Arabic sentence. Eventually we store the original Danish sentence and its translation into our DA-AR parallel corpus. We retrain the baseline again with the new bilingual data. We show that a parallel corpus obtained using this technique helps considerably to improve our SMT baseline. Sections 5.1 and 5.2 describe this process in details. We inspect another approach to improve our baseline performance. We mine available DA-EN and EN-AR language resources to extract parallel DA-AR sentences. We try to find all common English sentences between the two groups. Figure 1 explains this process. Once we have a match we extract the Danish sentence from the DA-EN part and the Arabic sentence from the EN-AR part. We store the new parallel extracted sentence pair into our DA-AR corpus. This step will help increase our parallel Danish-Arabic data size. To achieve that first we use Lemur to build indexes on our EN-AR resources. We start the search from the DA-EN side. We prepare a window size of four words and search all the possible occurrences of that window in the EN-AR side. If we find a match then we extract the Danish part that is parallel with the English window. We do the same with the match English sentence and extract the parallel Arabic part of that sentence. Lemur IR system is used to minimize the search space while searching for matches and will locate the most likely document that contains a matching English sentence at the EN-AR side. We apply the longest common subsequence problem approach to find a similar common string between resources. Section 5.3 explains more details of our approach. We retrain our baseline with the new extracted data, this will help boost the baseline performance. Finally we show that new parallel data extracted from that step which is not similar to the baseline training data domain will help improve the system performance, but not significantly.

> Published By: Blue Eyes Intelligence Engineering & Sciences Publication Pvt. Ltd.



Example 1 Example 1 Danish Source Hun svarede mig ikke og jeg fortrød med det samme, at jeg havde spurgt. English Translation She did not answer my question and I immediately regretted having asked Arabic Reference Query (System Jarien on aver of the optic optic of the optic of the optic of the optic optic	Table 2: Translation examples produced by our baseline				
Example 1 Danish Source Hun svarede mig ikke og jeg fortrød med det samme, at jeg havde spurgt. English Translation She did not answer my question and I immediately regretted having asked Arabic Reference She did not answer my question and I immediately regretted having asked Query (System Ja تجيئي وشعر ندم فورا بعد سوال Translation	and query r	esults returned by Lemur IR tool			
Danish Source Hun svarede mig ikke og jeg fortrød med det samme, at jeg havde spurgt. English Translation She did not answer my question and I immediately regretted having asked Arabic Reference Immediately regretted having asked Query (System Translation) Immediately regretted having asked Result 1 Lörzei, ender ver del units and the second second second second second second second med det samme, at jeg havde spurget Result 2 Danish Source Immediately regretted second sec		Example 1			
med det samme, at jeg havde spurgt. English Translation She did not answer my question and I immediately regretted having asked Arabic Reference Immediately regretted having asked Query (System Translation) Immediately regretted having asked Result 1 She did not answer my question and I med basic med usic Immediately regretted having asked med law Immediately regretted having asked Martin 2 Immediately regretted having asked med law sequence Martin 2 Immediately regretted having asked Martin 2 Immediately regretted having asked Kesult 3 Statistica static factor Martin 2 Immediately regretted having asked Martin 2 Immediately regretted having asked Martin 2 Statistica asked statica statistica Italization Immediately regretted having asked Beg er sikker på, at vi snart får dem at se igen, "sagde jeg, i et forsog på at glatte ud. I armslation I am sure we will see them again. "I statica statica statica statica statica Reference Daving back statica Query Statica batola datica statica M	Danish Source	Hun svarede mig ikke og jeg fortrød			
English Translation She did not answer my question and I immediately regretted having asked Arabic Reference لي تجبني وشعر ت بالندم مباشرة بعد ما سالته Reference Query (System Translation) لي تجبني وشعر ندم فورا بعد سوال مرتجيني , اقتريت اكثر فلم تجبني ، محددت La تجبني , اقتريت اكثر فلم تجبني ، محددت La تجبني , اقتريت اكثر فلم تجبني ، محددت La تجبني وامسكت بكتفها رسالتها La تجبني فد نهضت وهي تنظر بندم لأحد الطق مباشرة و لا حتى التعليل الطار لات خلفا، بعد ما القاه من كلام لم استط والطار لات خلفا، بعد ما القاه من كلام لم استط والعلو الات خلفا، بعد ما القاه من كلام لم استط والعلو الات خلفا، بعد ما القاه من كلام لم استط والعلو الات خلفا، بعد ما القاه من كلام لم استط والعلو الات خلفا، بعد ما القاه من كلام لم استط والتعلق مباشرة و لا حتى التعليل العلو الات عليل العلو الالي والي والي والتا سار مع مرة اخرى, لقد قلت ذلك العماد الام مرة اخرى, لقد قلت ذلك التعليف الامور والتا متكد النا متراهم راي قلى نذلك تحسين الاشياء والتا سنر اهم مرة اخرى, لقد قلت ذلك التما متكده الني مارك مرة اخرى, لقد قلت ذلك والتا سنر اهم مرة اخرى, لقد قلت ذلك التما متكده الي مرة اخرى, لقد قلت ذلك والتا سنر اهم مرة اخرى, لقد قلت ذلك التما متكده الي ماري الام مرة اخرى, لقد قلت ذلك والتا سنر اهم مرة اخرى, هذا راي والتا سنر اهم مرة اخرى هذا راي والتا سنر اهم مرة اخرى الخرى بهذا راي والتا سنر اهم المنافسة. ورانه هر المنوب والتا سنر اهم اخذت افكار يمنا منوب والتا سنر اهم اخرى الم مرة الجرى الم الم والوج والتا سنر اهم اخذت افكار يمنا من الم واحز الي المت والم الوم اخذت افكار ي منا المنافي الم والم واحز الي التي والت الم والم واحز الي التي والفران الم الم الم الموا المرب الم الم الم الم اخلي الي المي وال		med det samme, at jeg havde spurgt.			
Translation immediately regretted having asked Arabic له تجبني وشعر تالذم مباشرة بعد ما سالتها Query (System لم تجبني وشعر ندم فورا بعد سوّال Result 1 ده عدما على Nation 1 سوّالي Result 2 سوالي ي مرتجبني القريب الكلار فادت كلانه فادن و لم تعظير ندم لأحد سوالي Result 3 سوالي Result 3 بي تعظير ندم لأحد ي تو واسمكت بكنتها وسالتها سوالي الم تجبني فقد نهضت و هي تنظر ندم لأحد بي تعلير ندم لأحد الطار لات خلفنا، بعد ما آلفاه من كلام لم سنطع العلي ال الطرة و لا حتى التطيل العربي التها و التها من كلام لم سنطع Translation I am sure we will see them again. "I said that in an attempt to smooth things over" things over" Arabic والتنا سار هم مرة أخرى القد قلت ذلك Query والتا سار هم مرة أخرى و الم اليان ستصد. الم متكاد الذي قلد و بينه الم مالي تنصور و بينه Query والتا سار و الماليان ستصد. لم منالي و مالي مرد التي و بينه الم مالي و بينه Query والمالي الم مرد التي و بي في الله مرد و بي في اللله مرد و الحي و بي في اللله مرد الحي و اللي و بي في	English	She did not answer my question and I			
Arabic Reference سرات بالندم مباشرة بعد ما سالتها Query (System الم تجبنى وشعر ندم فورا بعد سزال Result 1 سوالى سوالى Result 2 الم تجبنى , اقتربت اكثر فاكثر ، ولم تجبنى ،مددت يدى و امسكت بكنها وسالتها Result 3 الم تجبنى , اقتربت اكثر فاكثر ، ولم تجبنى ،مددت الطو لات خلفنا، بعد ما اقاد من كلام/م استطح النطق عند نهضت و مى تنظر بندم لاح Danish Source Jeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud. English I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic الا am sure we will see them again. "I said that in an attempt to smooth things over" Query النا متأكد النا ساز اهم در أي قال ذلك تحسين الاشيا ولانا سنز اهم در أي قال ذلك تحسين الاشيا ولانا سنز اهم در أخرى, الذي ما تمزك ولانا شراه م درة اخرى, النه قاد ولايا سرارى الامل در الجى بر ألك ولايا سرارى الامل در الخي قال ذلك المور المارى الني معاد ولايا سرارى المار در الخي هذا راي ولايا سرارى الم مرة اخرى هذا راي ولايا متأكد باتى قاد على كسب الثقه و أنا لست ولايا منز الم م درة اخرى هذا راي ولايا منز الم مرة اخرى هذا راي ولايا منز الم مرة اخرى هذا راي ولايا للم مرة اخرى هذا راي ولاي منز الم مرة اخرى هذا العاد راي منا للم الم برو الم من الان اليوم اخدت القاري منا در السليم اخدت القاري منا در النيم الم والوم ذهبت افكار جديد بر الم حلي السليم الوم ذهبت افكار جديد بر الم حلي السليم الوم ذهبت افكار جديد بر الم حلي القدي م الور نم القدر نمبت الم من الان البذور الم نم الم والي الم الم و الم	Translation	immediately regretted having asked			
Reference الم تجبني وشعر ندم فورا بعد سوّال Query (System Translation) الم تجبني وشعر ندم فورا بعد سوّال Result 1 الم تجبني وشعر المدم مباشرة بعدها على سوّالي Result 2 الم تجبني وشعر بالالم مباشرة بعدها على الم تجبني وامسكت بكنفها وسالتها بدي وامسكت بكنفها وسالتها Result 3 الم تجبني وشعر بالالم المتلخ الطاو لات خلفا، بعد ما القاه من كلام لم استلخ الطاق مباشرة و لا حتى التعليل Danish Source Jeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud. English Translation I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic الالم مرة الحرى, الذ قلت ذلك thigs wer" Query النا متاكد النا ماري قال ذلك تحسين الأشيا Query النا متاكد النا مور الي قال ذلك تحسين الأشيا Query الم تحري إلي قال ذلك تحسين الأسي الذي ألم مرة الحرى, قالة والي المت والنا سراهم مرة الحرى, قالت ذلك المور بينه و بينه و اللها Result 1 الم تكد بالم مرة الحرى, قالة وال والنا سراهم مرة الحرى, ها والي ستصد. Result 3 الم تحري الم تحري الم تعري والنا سراهم مرة الحرى و الحرى و الجني والنا سراهم مرة الحرى و هالي و العالي سراهم و والنا سراهم مرة الحرى و هالي و الي المالي و و والنا سراهم مرة الحرى و هالي و المالي الحرى و والنا سراهم مرة الحرى و المور و للهاي و و المالي و المي العور و المالي و الحرى المور و المالي و المور المور و يو المالي و المالي و و مال الموزالي القور و لعرالي و الحرالي اللامي و و مالو مذهيت الم	Arabic	لم تجبني وشعرت بالندم مباشرة بعد ما سألتها			
Query (System Translation) فرا بعد سؤال الم تجبنى , أحسست بالندم مباشرة بعدها على سوالى الم تجبنى , اقتريت اكثر فلكثر ، ولم تحينى ،مددت العلي لات خلفا، بعد ما لقاه من كلام ام استطع الطلو لات خلفا، بعد ما لقاه من كلام ام استطع الطلو لات خلفا، بعد ما لقاه من كلام ام استطع الطلو لات خلفا، بعد ما لقاه من كلام ام استطع الطلو لات خلفا، بعد ما لقاه من كلام ام استطع والعلي الطلو لات خلفا، بعد ما لقاه من كلام ام استطع الطلو لات خلفا، بعد ما لقاه من كلام ام استطع الطلو لات خلفا، بعد ما لقاه من كلام ام استطع والتع مي القام والع حتى التعليل الطلو لات خلفا، بعد ما لقاه من كلام ام الطلو لات خلفا، بعد ما لقاه من كلام ام الطلو لات خلفا، بعد ما لقاه من كلام ام القام الم الم الم المتطع والتا متاكد النا سلز المع مرة أخرى, لقد قلت ذلك التا متاكد النا سلز اهم مرة أخرى, لقد قلت ذلك التا متاكد النا سلز اهم مرة أخرى, لقد قلت ذلك التا متاكد النا سلز اهم مرة أخرى, لقد قلت ذلك التا متاكد النا سلز اهم مرة أخرى, قلت قلت ذلك التا متاكد النا سلز اهم مرة أخرى, قلت قلت التا متاكد النا سلز اهم مرة أخرى, قلت قلت التا متاكد النا سلز اهم مرة أخرى, قلت قلت التا متاكد الن سلز ام مرة فرانا لله من التا متاكد الن سلز ام مرة فرانا لله ست التصين الامور بيني و بيني الامي الذي أنا متاكد منه اللبان ستصمد. Result 1 علي مرة أخرى, قلت قلت النا متاكد بلي قلار إلى مرة أخرى المت الام ماكد الي قلار على مرة أخرى المته والي الم مرة أخرى هذا راي الام من الذي أنا متاكد منه مالبان ستصمد. Result 3 قلت من الماني الم مرة أخرى هذا راي الم من النا منافر الم مرة أخرى هذا راي الم من الم مرة أخرى المالي منه الذي ألم من الم الم الم من النا البور أخب من المالي الم بوز ألي من الام الي م ذهت المالي من الم من الن الم الم من النا البنور ألمي من الام الو م أخلت أفكار مي من النا الم الم والم الم الم الم المعاد و الم	Reference				
Translation) Result 1 سوالي سائندم مباشر ة بعدها على الم تجبني , أختست بقذ نه مباشر ة بعدها على الم تجبني , أفتر بت أكثر فأكثر ، ولم تجبني ، مددت 2 Result 2 نعو إمسكت بكتها وسائتها ، مددت 2 يدي و إمسكت بكتها وسائتها ، مددت 3 يدي و إمسكت بكتها وسائتها ، مددت 3 Result 3 نع أقاد من كلام إمستط ع الطول لات خلفنا، بعد ما أقاد من كلام إمستط ع يدي و إمسكت بكتها وسائتها ، مدد 1 Danish Source Jeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud. English I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic I am sure we will see them again. "I said that in an attempt to smooth things over" Query الم مرة أخرى إقد قلك ذلك تعسين الاشيا = واننا ستراهم رأي قل ذلك تعسين الاشيا = Result 1 ظلف ذلك مرة أخرى إقد ذلك الم المتكاد نو سازى الأمر مرة أخرى إقد قلك ألما من الأمر أي قاد علك ألما منا الأمر أي قاد إلى الم مرة أخرى إقد ألك ألما متكاد نو سازى الأمر مرة أخرى إقد ذلك ألم ألما الم الم مرة أخرى إلم فلك فلك ألما من ألمر أي ألم ألم مرة أخرى إلم فلك فلك ألما متكاد ألم مرة أخرى إلم فلك فلك ألما متكاد ألما مراهم أي قاد على مرة أخرى إلم فلك فلك ألما ألما الم عدى ألم مرة أخرى إلم مرة أخرى إلم فلك فلك ألما ألما الما مدى ألم مرة أخرى إلم فلك فلك ألما ألما معال إلما فلك ألما ألما مك فلك ألما ألما ألما ألما مدة ألما ألما ألما مدة ألما ألما ألما مدة ألما ألما ألما ألما ألما ألما ألما مدى ألم من الأل ألما ألما ألما مدة ألما ألما مدى ألما ألما مدين ألما ألما ألما مدى ألما ألما ألما ألما ألما ألما ألما ألما	Query (System	لم تجبني وشعر ندم فورا بعد سؤال			
Result 1 سؤالی سؤالی سؤالی Result 2 نعند، ولم تجبنی مددت سؤالی سؤالی Result 3 نعند، ولم تجبنی هذ نهضت و می تنظر بند لأحد الطاو لات خلفنا، بعد ما ألقاه من كلام لم استطع Danish Source Jeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud. English I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic ظلف تلف تلف تلف تلف الم استطع Query قائ تلف تحسين الام ور اي قال ذلك تحسين الاشيا Query قائ متأكد سنر اهم مر أي قال ذلك تحسين الاشيا Result 1 قائ متأكد من الم مراي قال ذلك تحسين الام ور بيني و بينه النا متأكد الن الم مراي قال ذلك تحسين الاشيا قائ متأكد من الم مراي قال ذلك تحسين الام ور بيني و بينه Result 2 قائ متأكد من الم مراي قال ذلك تحسين الاشيا Result 3 قائ متأكد من الم مراي قائ مأكد من القاه وراي لي مت وائنا متأكد من الم مراي قائ مأكد من القاه ورا لي مت Result 3 قائ مالي مع مراي فر فر الي مت الم مرا الم ور الي مت قائ اللي مندور الي مراي ور الته مي ور الي مي Result 3 قائ مالي مراو الي من الم مرا ور اي قائ مت آلف من ذلك الي مي ور اي قائ مالي مراي ور الم مراي ور الي مي ور الي مي ور الي مي	Translation)				
سواليResult 2سرتيبني اقتريت اكثر فاكثر، ولم تجنيني ،مددتkresult 3ديدي وامسكت بكتفها وسالتهاkresult 3ديدي وامسكت بكتفها وسالتهاKesult 3ديدي وامسكت بكتفها وسالتهاKesult 3ديدي وامسكت بكتفها وسالتهاKesult 3ديدي وامسكت بكتفها وسالتهاBarbon 1Example 2Danish SourceJeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud.English TranslationI am sure we will see them again. "I said that in an attempt to smooth things over"Arabic Queryالقلم الخرى, القد قلت ذلك الا متاكد انا سرا هم رة اخرى, لقد قلت ذلك الا متاكد انا سرا هم رة اخرى, وقلت ذلك الا متاكد الني المراح مرة اخرى, وقلت ذلك الا متاكد الني المراح ولي قال ذلك تحسين الاثنياResult 1 Queryديني و بينه الا متاكد الني المراح ولي قال ذلك تحسين الاثنيا واننا سرا هم رة اخرى, وقال ذلك وانا المراح والي قال عراح مرة اخرى الاثنيا معاد الني والم مرة اخرى, وقال ذلك والنا سنر اهم مرة اخرى, وفار الي والنا سنر اهم مرة اخرى والي المراح والي والي والنا سرا هم مرة اخرى والي الحراح والنا سنر اهم مرة اخرى والي والي والنا سرا مم مرة اخرى والي المراح والنا سرا مم مرة اخرى والي الح والنا سنر اهم مرة اخرى والي والي والنا سنر اهم مرة اخرى والي والي المراح والنا مين والي والي والي والي المراح والي المراح والي والي المراح والي والي والي والي والي والي والي والي	Result 1	<i>لم تجبني ,</i> أحسست بالندم مباشرة بعدها على			
Result 2 م.مددت يدي واسمكت بكتها وسالتها لايت و المسكت بكتها وسالتها النطق مباشرة و لا حتى التعليل الطاو لات خلفنا، <i>بعد ما</i> القاه من كلام استطع النطق مباشرة و لا حتى التعليل Result 3 ع كلام استطع الطاو لات خلفنا، <i>بعد ما</i> القاه من كلام استطع النطق مباشرة و لا حتى التعليل Danish Source Jeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud. English Translation I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic انا متاكد النا منر اهم مرة اخرى, لقد قلت ذلك انا متاكد الني مرام مرة اخرى, القد قلت ذلك انا متاكد الني مرام مرة اخرى, هذا لاتيا. Result 1 طلق تحصين الأشيا ، معاد المرام مرة اخرى, هذا راي وائنا سنر اهم مرة اخرى, هذا راي الكلي فالي مراري والمل مرة اخرى هذا راي وائنا سنر اهم وا ال سات وائنا سنر اهم وا ال سات العداد والم بونو الني يتثنيها ومنذ ذلك اليوم اخذت فكار على كست القه و العداد الجر اني القد والم الما ور وائنا سنر المو والي مات وائنا سنر المو اخذت الجر الي القد والم ومنذ ذلك اليوم ذهبت افكار جديدة بدل عديق احران العراني القد ما فنال الما وا ال سابلي ومنذ ذلك اليوم اهي الما و اخر الي القد الجر الي ومنذ ذلك اليوم أهي ال الخر الي القد الي العراني القدا العو الي الني القد الي العراني القد الجر الي القد الجر الي القد الي والغر الجر الي القد الي القد الي المي الي والغر الجر الي القد الي المي الخر الي القدام والمن الذو الجر الي المي الجر الي المي الجر ال		سؤالي			
الجنوب الجنوب المرتجبني فقد نهضت و هي تنظر بندم لأحد الطاولات خلفنا، بعد ما ألقاه من كلام لم استطع السلولات خلفنا، بعد ما ألقاه من كلام لم استطع العالي الطاولات المعالي العام المعالي العام المعالي العام المعالي العام المعاليExample 2 Danish SourceJeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud.English TranslationI am sure we will see them again. "I said that in an attempt to smooth things over"Arabic ReferenceI am sure we will see them again. "I said that in an attempt to smooth things over"Queryالتامية الذي آنا متاكد النا سنر اهم مرة اخرى, لقد قلت ذلك العام المي الأمر المي مرة اخرى, قلت ذلك العام المعار الم مرة اخرى, هذا راي عالي المي عالي الأمر المي مرة اخرى, هذا راي عالي المي عالي الأمر الي علي الأمر التي المي الم مرة اخرى, هذا راي عالي المي عالي الأمر الم مرة اخرى, هذا راي عالي عالي معالي الأمر الله المعار الأمر الم مرة اخرى القدي أنا متاكد مند الم مرة اخرى, هذا راي عالي المعاد الم مرة اخرى, هذا راي عالي عالي معالي الأمر الي المعاد عالي معاد الله والي ما تكل الأمر الم مرة اخرى الأمر الإمراع عالي معاد المعاد العام المعاد عاد المعاد الم مرة اخرى الألي المعاد عاد المعاد المعاد المعاد عاد المعاد المعاد عاد المعاد الم مرة اخرى الأمر الأمر الأمر الأمر الأمر الأمر ما أور الأمر المعاد عاد المعاد المعاد المعاد الألي الأمر الأمر الأمر الأمر المعاد عاد المعاد المعاد عاد الله والأمر المعاد المعاد عاد المعاد عاد المعاد عاد المعاد المعاد المعاد عاد المعاد عا	Result 2	<i>لم تجبني</i> اقتربت اكثر فاكثر، ولم <i>تجبني</i> ،مددت			
Result 3 الم تجنبنى فقد نهضت و هي تنظر بندم لأحد الطاو لات خلفانا، بعد ما ألقاه من كلام لم استطع الطاو لات خلفانا، بعد ما ألقاه من كلام لم استطع اللطاق برات خلفانا، بعد ما ألقاه من كلام لم استطع اللطاق برات خلفانا، بعد ما ألقاه من كلام لم استطع الطاو لات خلفانا، بعد ما ألقاه من كلام لم استطع العلي العلوي التعلوي العلوي العلو		يدي وامسكت بكتفها وسألتها			
الطار لات خلفنا، <i>بعد ما</i> القاه من كلام/لم استطع النطق مباشرة و لا حتى التعليل العلق مباشرة و لا حتى التعليل Example 2 Danish Source Jeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud. English Translation I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic Query التا متأكد انا مر أم قال ذلك تحسين الأشياء انا متأكد اني مراح ارفي قال ذلك تحسين الأشياء الما متأكد سنرا هم مرة اخرى, هذا راي واتنا سنرا هم مرة اخرى, هذا راي واتنا سنرا هم مرة اخرى, هذا راي واتنا سنرا هم مرة اخرى, هذا لي الثاري واتنا سنرا هم مرة اخرى, هذا لي الثاري واتنا سنرا هم مرة اخرى, هذا لي الثاري واتنا سنرا هم مرة اخرى هذا لي الإلي واتنا سنرا هم مرة اخرى هذا لي واتني واتنا سنرا هم المنافي واتنا سنرا هم المي والما مي واتنا سنرا هم المنافي واتنا سنرا هم المي والم المالي والم المولود والفي والغار الجرية العربي واتا البوم اخذت القاري منحا جدية العربي واتا البوم اخذت القاري منحا جدية العربي واتا البوم اخذت القاري منحا جدية العربي والفي الي مذلك الجرية والي مي واتا السبحث أفر الحاني التني والي والي والي التنيني على المسير والسابي والي الخراني التي والوعار المي الخران الجراني التي والوعار المي والي الحراني التي والغار الجرية والرضا, هم المي والي الي الإلي والغار الجراني الي منا الحراني العربي من والعار الحراني والمي الحراني الي مي والغار الجراني و	Result 3	لم تجبني فقد نهضت وٍهي تنظر بندم لأحد			
النطق مباشرة و لا حتى التعليل Example 2 Danish Source Jeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud. English I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic Jain attempt to smooth things over" Arabic Baitb E Tenglish that in an attempt to smooth things over" Reference Junto and E textor, the after and the addition a		الطاولات خلفنا، <i>بعد ما</i> ألقاه من كلام <i>لم</i> استطع			
Example 2Danish SourceJeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud.English TranslationI am sure we will see them again. "I said that in an attempt to smooth things over"Arabic ReferenceI am sure we will see them again. "I said that in an attempt to smooth things over"Queryدان متاكد اننا سنز اهم مرة اخرى, القد قلت ذلك انا متاكد انن الشيء الذي <i>أنا متاكد منه أن طالبان ستصد.</i> Result 1Essuit 1 الا متاكد الشيء الذي <i>أنا متأكد منه أن طالبان ستصد.</i> Result 2 دانا سنز اهم مرة اخرى, هذا ر اي و النا سنز اهم مرة اخرى, هذا ر اي و النا سنز اهم مرة اخرى هذا ر اي و النا سنز اهم مرة اخرى و بينهResult 3Example 3Danish SourceFra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg.English TranslationFrom that day, my thoughts went to new directions, instead of my old deep sorrow.Arabic new directions, instead of my old deep a oti Like a latic li القديمة a oti Like a latic li القديمة a oti Like a set. عدية الحالي النا اللاي a tiel li المتي منا لي منا لي منا ي م منا ي منا ي منا ي منا ي م منا ي منا ي م م الان الم الم ي م م الان الم م م الان الم م م الان م م الان م م الان م م م الان الم م م الان الم م م الان م م م الان الم م م الان الم م م الان م م م الان الم م م الان الم م م الان الم م م الان م م م الان الم م م الان الم م م الان الم		النطق مباشرة ولاحتى التعليل			
Example 2Danish SourceJeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud.English TranslationI am sure we will see them again. "I said that in an attempt to smooth things over"Arabic ReferenceI am sure we will see them again. "I said that in an attempt to smooth things over"Queryداندا متاكد انذا سنر اهم رأي قال ذلك تحسين الأشياءResult 1It arabic is an attempt to smooth things over"Result 2داندا متاكد منه أن طالبان ستصمد.Result 3Example 3Result 4Example 3Danish SourceFra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg.English TranslationFrom that day, my thoughts went to new directions, instead of my old deep sorrow.Arabic that 2Yer and jer and j					
Danish SourceJeg er sikker på, at vi snart får dem at se igen," sagde jeg, i et forsøg på at glatte ud.English TranslationI am sure we will see them again. "I said that in an attempt to smooth things over"Arabic ReferenceI am sure we will see them again. "I said that in an attempt to smooth things over"Queryالنا متاكد النا سن(ha o, ci الحرى, لقد قلت ذلك Result 1Result 1I am sure we will see them again.Result 2قلت كان الشيء الذي <i>أنا متأكد من</i> ه أن طالبان ستصمد.Result 3Example 3Result 4Example 3Danish SourceFra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg.English TranslationFrom that day, my thoughts went to new directions, instead of my old deep sorrow.Arabic (DueryYat, yat, act also a jo fine direction direc		Example 2			
se igen," sagde jeg, i et forsøg på at glatte ud. English Translation I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic Said that in an attempt to smooth things over" Arabic Said that in an attempt to smooth things over" Query Etadab Na or i Eda S, jab S Iti aritZe mit (Ba of ja Sit Lib Senut) Iti aritZe mit (Ba of ja Sit Senut) Result 1 Said Senut) Iti aritZe mit (Ba of ja Sit Senut) Said Senut) Result 2 Said Iti aritZe mit (Jago nei Senut) Result 3 Said Iti man. <i>full mit (Ba of Senut)</i> Said Iti man. Result 3 Said Iti man. <i>full mit (Ba of Senut)</i> Said Iti man. Result 3 Said Iti man. <i>full mit (Ba of Senut)</i> Said Iti Man. Result 3 Fra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg. From that day, my thoughts went to new directions, instead of my old deep sort exit i bevæga sut stall verke ja verke j	Danish Source	Jeg er sikker på, at vi snart får dem at			
English Translation I am sure we will see them again. "I said that in an attempt to smooth things over" Arabic Reference I am sure we will see them again. "I said that in an attempt to smooth things over" Query فلاف الأمور I arible الذي أنا متأكد منذ أن طالبان ساحمد. Result 1 فلاف الأمور فلاف الأمور فلاف الأمور Result 2 فلاف الأمور Result 3 فلاف الحرى, هذا راي واننا سنراهم مرة الحرى هذا راي وانا سنراهم مرة الحرى والنا سنراهم والم الحراني والم الحراني والم الحراني والم المراجع الم الحراني والم الحراني ومنذ ذلك اليوم الحزاني القديمة ومنذ ذلك اليوم الحزاني القديمة ومنذ ذلك اليوم الحزاني القديمة والما حديدة الكار والحدية اليوم الحزاني القديمة والما والم الحزاني الم الحراني والم الحراني والنا اليوم أهذت الكار والحدية اليوم ما حذات الماري والماري الحراني الم الحراني عنا المالي والنا المالق المالي الماليور ألم الحزاني القديمة والماليوا الحراني الم الحدية اليوم ألم الحراني والماليوا الماليو ملي الماليور والحياني الماليور والمار الحراني والماليو مي العمية بالمساعر والساليو الحراني فالمال الماليور والم الماليور الحراني والمال الحديدة والى من الان الدؤوا في زرع الكار الماليو والمار الحران الحران والما الحران والما الماليو الحراني والحالي الماليور والما الماليو الحران والما الحران والحران والما الحراني والمالمالجديدة اللما الحران والما الما الحديذ الكا		se igen " sagde jeg i et forsøg på at			
English TranslationI am sure we will see them again. "I said that in an attempt to smooth things over"Arabic Referenceالالم الم الم الم الم الم الم الم الم ال		glatta ud			
English TranslationI am sure we will see them again. I said that in an attempt to smooth things over"Arabic ReferenceSaid that in an attempt to smooth things over"Queryالنا متاكد الذا مرة اخرى, قلت ذلك الا متاكدة الي سارى الامل مرة اخرى, قلت ذلك الامير بيني و بينهResult 1الامير الإمير و ينه الامير بيني و بينهResult 2الكن الشيء الذي/أنا متأكد منه أن طالبان ستصمد. التحسين الامور بيني و بينه الله مرة اخرى, هذا راي العالي المتصعد.Result 3الامير الذي متأكد منه أن طالبان ستصمد. المي مرة اخرى, هذا راي الله مرة اخرى هذا راي الله مرة اخرى هذا راي الله مراح الذي متأكد منه أن طالبان ستصمد.BasisExample 3Danish SourceFra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg.English TranslationFrom that day, my thoughts went to new directions, instead of my old deep sorrow.Arabic ReferenceSourceQuery out ikies i bell ي util a ot icit ibil و مند ذلك اللي منا و مند ذلك الله اليوم الحدت الفكار عنا منا و مند ذلك الألا مراحي منا المناهم أن منا و مند ذلك الله مراحي منا و مند الما و منا و مند الما و مند الما و منا و مند الله اليوم الحدا الما و منا الماليوم الحدا الله المالي المالي المالي الماليوم الحدا المالي المالي المالي منا و مالي مالي مالي مالي و مالي مالي و مالي مالي و مالي مالي و مالي مالي و مالي و مالي و مالي و مالي و مالي مالي و مالي مالي و مالي و مالي و مالي و مالي و مالي و م	English	glatte ud.			
I ransiationsaid that if a fatchip to shooth things over"Arabic Referenceفي الأمور النا متاكد الذي الله مرة الحرى, لقد قلت ذلك الالمي الممر رأي قال ذلك تحسين الأشياءQuery النا متاكد الني سارى الأمل مرة الحرى, قلت ذلك الالمي المي رافي من المل مرة الحرى, قلت ذلك الالمي المي رافي من الملبان ستصمد.Result 1 النا متاكدة الي سارى الأمل مرة الحرى, قلت ذلك النا متاكدة الي سارى الأمل مرة الحرى, قلت ذلك الالمي الذي <i>أنا متأكد</i> منه أن طالبان ستصمد.Result 2 النا منز أهم مرة الحرى, هذا راي النا متأكد بأني قادر على كمب الثقه و <i>أنا</i> لست العنا منز أهم مرة الحرى, هذا راي النا متأكد بأني قادر على كمب الثقه و <i>أنا</i> لست العام مرة الحرى إلى هذا راي القد من المنافسه . وانه سيغيرون رائيهم العام علاء العام المنافسه . وانهم سيغيرون رائيهم العام مرة الحرى إلى المت العام مرة الحرى هذا راي العام مرة الحرى هذا راي العام مرة الحرى هذا راي العام مرة المي منافع منافع المي المي المي المي المي المي المي المي	English	I am sure we will see them again. I			
Arabic Referenceالنا متاكد انذا سنر اهم مرة اخرى, لقد قلت ذلك النا متاكد سنر اهم رأي قال ذلك تحسين الاشياء النا متاكد شاهم رأي قال ذلك تحسين الاشياء النا متاكد شاهم رأي قال ذلك تحسين الاشياء الامر راي قال ذلك تحسين الاشياء الامر راي قال ذلك تحسين الاشياء الامر راي قال ذلك تحسين الاشياء المحمدة اذي سارى الامل مرة اخرى, قلت ذلك المحمدة اذي سارى الامل مرة اخرى, قلت ذلك التحسين الامور بيني و بينه العن الشيء الذي أنا متأكد منه أن طالبان ستصمد.Result 1 المنافسيء الذي أنا متأكد منه أن طالبان ستصمد.Result 2 الما مرة اخرى, هذا راي العن المتاكد منه أن طالبان ستصمد.Result 3المنافسية الذي أنا متأكد منه أن طالبان ستصمد.Result 3المنافسية الذي أنا متأكد منه أن طالبان ستصمد.Result 3الما متأكد باتي قادر على كسب الثقه و أنا لست بسر عه في علاءResult 3Danish SourceFra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg.English TranslationFrom that day, my thoughts went to new directions, instead of my old deep sorrow.Queryالمان إليوم اخذت افكاري منحا جديدا , بدلا عن احزاني القديمةResult 1من اليوم ذلك اليوم يمنح القدية بدل عميق احز ان الماليةResult 1من اليوم ذهبت افكار جديدة بدل عميق احز ان السابق , ان احزاني القديمةResult 1المالية الماليةالماليةالمالة الماليةResult 2المالية الماليةالماليةالمالية الماليةالماليةالمالية الماليةالماليةالمالية الماليةالماليةالمالية الماليةالماليةالمالية الماليةالماليةالمالية الماليةالماليةالمالي	Iranslation	said that in an attempt to smooth			
Arabic العادلة العاسر أهم مرة الحرى, لعد قلت ذلك Reference النا متاكد سنر أهم رأي قال ذلك تحسين الأشياء النا متاكد آلي سارى الأمل مرة اخرى, قلت ذلك Result 1 العمل مرة اخرى, هذا راي المعني الأمور بيني و بينه Result 2 المعني الأمل مرة اخرى, هذا راي المعني الأمل مرة اخرى, هذا راي Result 3 المعني مرة اخرى, هذا راي المعني مرة اخرى, هذا راي Result 3 المعني مرة اخرى, هذا راي المعني مرة اخرى, هذا راي المعني مرة اخرى, هذا راي المعني معني مرابع مرة اخرى, هذا راي المعني معني مرابع مرابع مرابع مروز المع مرة اخرى منا المنافسه , وانهم مرة اخرى هذا راي المعني معني مروز رائي متاكد ماني مرابع مروز رائي معني مروز رائيهم المية المي مروز المع مروز راي المعني معني معني مروز رائيهم مروز مرابع مروز رائيهم سيغيرون رائيهم مروز رائيهم مروز مرابع مروز مرابع مروز رائيهم مروز مرابع مروز رائيهم سيغيرون رائيهم مروز مرابع مروز مرا	A 1. *				
Reference العلمية المؤرر Query انا متاكد سنر اهم رأي قال ذلك تحسين الأشياء Result 1 النا متاكد الله مرة اخرى , قلت ذلك Result 2 المل مرة اخرى , هذا راي Result 3 واننا سنراهم مرة اخرى, هذا راي Result 3 واننا سنراهم مرة اخرى, هذا راي Result 3 السابة وانهم سيغيرون رائيهم Result 3 تاك ماتكد باني قادر على كسب الثقه و أنا لست Result 3 تالنا منراهم مرة اخرى هذا راي Result 3 تالنا معنار ما مرة اخرى هذا راي Result 3 تالنا معنار ما ما ما فلابان ستصمد. Result 4 السابة ما ما فلابان ستصمد. Result 5 تالما ما فلابان ستصمد. Result 6 Fra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg. From that day, my thoughts went to new directions, instead of my old deep sorrow. sorrow. Arabic Yene yene Query ما خاذك الانسان ما الور ذهبت افكار جديدة بدل عميق احزان ما اليوم ذهبت افكار جديدة بدل عميق احزان ما الور فلابا ساباق والو ما فلابا ساباق والولابا والولابا فلا فكار الحران Result 1 السابق والولابا والولابا فلا فكار الحران الما الحزان	Arabic	انا مناجد النا سنر أهم مرة أخرى, نقد قلب ذلك			
Othery العليان المراجع والي عان للعا للعليان المراجع والي عان العاري الأسليم المرة الحرى , قلت ذلك الشيء الذي أنا متأكد منه أن طالبان ستصمد. Result 2 الكن الشيء الذي أنا متأكد منه أن طالبان ستصمد. Result 3 واننا سنراهم مرة اخرى, هذا راي واننا سنت مرة اخرى, هذا راي واننا سنراهم مرة اخرى, هذا راي وانا لست منابع وانا للذي وانا ست مرة وانا لست من الان المراه وانا للنا اللهم مرة اخرى منحا جديد إر النا للناهم وانا ما وانا الناهم وانا وانا وانا وانا وانا وانا وانا وان	Reference	التلبية الأمور			
Result 1 ال المحادة التي المرافر بيني و بينه التحسين الأمور بيني و بينه التحسين الأمور بيني و بينه الكن الشيء الذي <i>أنا متأكد</i> منه أن طالبان ستصمد. Result 2 الكن الشيء الذي <i>أنا متأكد</i> منه أن طالبان ستصمد. Result 3 واننا سنراهم مرة اخرى, هذا راي واني الست Result 3 الحامة مرة اخرى, هذا راي واني الست Result 3 الحامة مرة اخرى, هذا راي واني الست Result 3 الحامة مرة اخرى, هذا راي واني الست Result 4 الحامة مرة اخرى, هذا راي واني الست Result 5 الحامة مرة اخرى, هذا راي واني واني واني الست Result 5 الحامة مرة اخرى, هذا راي واني واني واني واني واني واني واني وا	Query	الالمناحد سنراهم راي قال دلك تحسين الإسياء			
Result 2 الخاصين الامور بيلي و بيله و بيله الذي أذا متأكد منه أن طالبان ستصمد. واننا سنر اهم مرة اخرى, هذا راي واننا سنر اهم مرة اخرى, هذا راي Result 3 أنــــــــــــــــــــــــــــــــــــ	Kesuit I	الأملكدة التي تسارى الأمل مرة الحرى بالتك ذلك			
Result 2 العلى الشيء الذي الذي ماذ لمك ال طالبان النظمة. واننا سنراهم مرة اخرى, هذا راي انــــــــــــــــــــــــــــــــــــ	D	المحسيل الأمور بيني و بينه			
Result 3 والع السرافيم مرده الحري, هذا راي انسا متأكد بأني قادر على كسب الثقه و أنا لست خانف من المنافسه . وانهم سيغيرون رائيهم بسر عه في علاء جسر عه في علاء Danish Source Fra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg. English Translation From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic Reference From that day, my thoughts went to new directions, instead of my old deep sorrow. Query منا ليوم ذهبت افكار جديدة بدل عميق احزان العربية المسابق , ان احزاني القديمة المسابق , ان احزاني لن تثنيني عن المسير العماري Result 1 السابق , ان احزاني لن تثنيني عن المسير العماري Result 2 من الإن ابدؤوا في زرع افكار السلام الجديدة و المعارية العربي من الإن ابدؤوا في زرع افكار السلام العوز ان الاحزان Result 3 من الإن ابدؤوا في زرع افكار السلام الجديدة و المعارية المسابق , و احساس عميق بالمساور اليه الاحزان	Result 2	لكن اللاتيء الذي أن مناقد منه أن طالبان ستصمد.			
Kesult 3 السائة اللغة و ال السك خانف من المنافسه . وانهم سيغيرون رائيهم خانف من المنافسه . وانهم سيغيرون رائيهم بسر عه في علاء Example 3 Danish Source Fra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg. English From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic Sorrow. Reference عن احز اني القديمة Query نايوم ذهبت افكار جديدة بدل عميق احز ان السابق , ان احز اني لن تثنيني عن المسير من اليوم ذهبت الغوم بي نايوم في الانسان Result 1 السابق , ان احز اني لن تثنيني عن المسير السابق , ان احز اني لن تثنيني عن المسير و منذ ذلك الانسان Result 2 السابق , ان احز اني لن تثنيني عن المسير العميقة بالمشاعر و السابق , ان احز اني لن تثنيني عن المسير السابق , ان احز اني لن تثنيني عن المساعر و السابق , ان احز ان العميزة بالمشاعر و الماعميق بالمشاعر و المنابق ابذوا في زر ع افكار السلم الجديدة و العميق بالمساعر و من الان ابدؤوا في زر ع افكار السلم الجديدة و السابق , الاخز ان الاحز ان الاحز ان الاحز ان الماعروا في زر ع افكار السلم الجديدة و الاحز ان ال	D	والت للنظر العلم عمرة الحربي هذا راي			
Example 3 Danish Source Fra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg. English From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic Value at a ference Query Jule at a ference Danish 1 Norder at a ference Result 1 Norder at a ference Result 1 Norder at a ference Norder at a ference Norder at a ference Result 1 Norder at a ference Norder at a ference Norder at a ference Result 1 Norder at a ference Norder at a ference Result 2 Norder at a ference Norder at a ference Norder at a ference Result 3 Norder at a ference Norder at a ference Norder at a ference Norder at a ference Result 3 Norder at a ference Norder at a ference Norder at a ference Norder at a ference Result 3 Norder at a ference Norder at a ference Norder at a ference Norder at a ference Result 3 Norder at a ference Norder at a ference </th <th>Kesuit 3</th> <th>الت متاكد بالي قادر على كسب اللغة و ال لسب</th>	Kesuit 3	الت متاكد بالي قادر على كسب اللغة و ال لسب			
السر عدي عرب المرابعة في المرابعة في المرابعة في عرب المرابعة في عرب المرابعة في المرابعة ألية ألية في المرابعة ألية ألية ألية ألية ألية ألية ألية ألي		حاف من المنافسة ، <i>واليهم سيعلير ول راليهم</i>			
Example 3Danish SourceFra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg.English TranslationFrom that day, my thoughts went to new directions, instead of my old deep sorrow.Arabic ReferenceEver en		بشرعة في عارع			
Danish Source Fra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg. English From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic Yarabic sorrow. Reference aviet i lizuañ el ci lizuañ		Example 2			
Danish Source Fra da af begyndte mine tanker at bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg. English From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic Varie et in the sorrow. Query Varie et in the sorrow. Query Not like et in the sorrow. Result 1 Not like et in the sorrow. Result 2 Not like et in the sorrow. Result 3 Not like et in the sorrow. Result 4 Not like et in the sorrow. Result 3 Not like et in the sorrow. Not like et in the sorrow. Not like et in	Danich Source				
bevæge sig i helt nye retninger, hvoraf den første var en dyb sorg. English Translation From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic بالالله الله الله الله من الله الله الله الله الله الله الله الل	Damsii Source	Fra da af begyndte mine tanker at			
den første var en dyb sorg.English TranslationFrom that day, my thoughts went to new directions, instead of my old deep sorrow.Arabic ReferenceKitter (Karabic) sorrow.Queryناحزاني القديمةQueryالعربة الحزان السابق, ان احزاني لن تثنيني عن المسير المسيرResult 1السابق, ان احزاني لن تثنيني عن المسير المسيرResult 2المسير المسيرResult 3الاخلان العندة و الرضا, فعندما يحدث ذلك عليكم ان العدية و الرضا, فعندما يحدث المسير المسيرResult 3الاخلان العندة و الرضا, فعندما يحدث ذلك عليكم ان المسيرResult 3المسير العندة و الرضا, فعندما يحدث ذلك عليكم ان السعادة و الرضا, فعندما يحدث ذلك عليكم ان السعادة و الرضا, فعندما يحدث ذلك عليكم ان		bevæge sig i helt nye retninger, hvoraf			
English Translation From that day, my thoughts went to new directions, instead of my old deep sorrow. Arabic Reference Sorrow. Query الالتانية التالية المسبحت افكر باسلوب جديد ، بد ل عميق احزان المسابق , ان احزاني لن تثنيني عن المسير Result 1 السابق , ان احزاني لن تثنيني عن المسير Result 2 المسير Result 3 الاخكار الجديدة و الرضا , فعندما يحديث ذلك عليكم ان Result 3 الاختار الحديث المسير Result 3 الالان ابدؤوا في زرع /فكار السلام الجديدة و Result 3 السعادة و الرضا , فعندما يحدث ذلك عليكم ان تتكلموا بحزم و احساس عميق بالمسؤالية		den første var en dyb sorg.			
Translation new directions, instead of my old deep sorrow. Arabic sorrow. eait ذلك اليوم اخذت افكاري منحا جديدا , بدلا ومنذ ذلك اليوم اخذت افكاري منحا جديدا , بدلا Reference auril a control and	English	From that day, my thoughts went to			
sorrow. Arabic e aić ذلك اليوم اخذت افكاري منحا جديدا , بدلا عن احزاني القديمة Reference عن احزان القديمة avi اليوم ذهبت افكار جديدة بدل عميق احزان Query احزان noi اليوم ذهبت افكار جديدة بدل عميق احزان Result 1 السابق , ان احزاني لن تثنيني عن المسير السابق , ان احزاني لن تثنيني عن المسير الافكار الجديدة اليوم هي العميقة بالمشاعر و Result 2 الافكار الجديدة اليوم هي العميقة بالمشاعر و Result 3 الافكار الجديدة و في زرع <i>افكار</i> السلام الجديدة و تالافكار الجديدة اليوم هي العميقة بالمشاعر و الافكار الجديدة اليوم هي المسير تالافكار الجديدة اليوم و مي العميقة بالمشاعر و تالافكار الجديدة اليوم في زرع <i>افكار</i> السلام الجديدة و تالافكار الجديدة اليوم و المالي في زرع الحكار السلام الجديدة و الرضا , فعندما يحدث ذلك عليكم ان تاكلموا بحزم و احساس عميق بالمسؤالية	Translation	new directions, instead of my old deep			
Arabic Reference لابل اليوم اخذت افكاري منحا جديدا , بدلا عن احزاني القديمة Query من اليوم ذهبت افكار جديدة بدل عميق احزان Result 1 اصبحت افكرباسلوب جديد , لم اعد ذلك الانسان Result 1 السابق , ان احزاني لن تثنيني عن المسير السابق , ان احزاني لن تثنيني عن المسير Result 2 المسير الافكار الجديدة اليوم هي العميقة بالمشاعر و Result 3 الافكار الحديدة اليوم في زرع <i>افكار</i> السلام الجديدة و السعادة و الرضا , فعندما يحدث ذلك عليكم ان تتكلموا بحزم و احساس عميق بالمسؤالية		sorrow.			
Reference عن احزاني القديمة من اليوم ذهبت افكار جديدة بدل عميق احزان Query من اليوم ذهبت افكار جديدة بدل عميق احزان اصبحت افكرباسلوب جديد لعميق احزان Result 1 السابق ان حازان لن تثنيني عن المسير السابق ان احزاني لن تثنيني عن المسير المسير Result 2 الافكار الجديدة اليوم هي العميقة بالمشاعر و الافكار الجديدة اليوم هي العميقة بالمشاعر و الاحزان الاحزان العديدة و الرضا فعندما يحدث ذلك عليكم ان تتكلموا بحزم و احساس عميق بالمسؤالية السوالية	Arabic	ومنذ ذلك اليوم اخذت افكاري منحا جديدا , بدلا			
من اليوم ذهبت افكار جديدة بدل عميق احزان Result 1 اصبحت افكرباسلوب جديد , لم اعد ذلك الانسان Result 1 السابق , ان احزاني لن تثنيني عن المسير الافكار الجديدة اليوم هي العميقة بالمشاعر و الافكار الجديدة اليوم في ارع/فكار السلام الجديدة و من الان ابدؤوا في زرع/فكار السلام الجديدة و السعادة و الرضا , فعندما يحدث ذلك عليكم ان تتكلموا بحزم و احساس عميق بالمسؤالية	Reference	عن احز اني القديمة			
Result 1 اصبحت أفكر باسلوب جديد , لم اعد ذلك الانسان السابق , ان احز اني لن تثنيني عن المسير الافكار الجديدة اليوم هي العميقة بالمشاعر و Result 2 الافكار الجديدة اليوم هي العميقة بالمشاعر و الافكار الجديدة اليوم هي العميقة بالمشاعر و من الان ابدؤوا في زرع/فكار السلام الجديدة و السعادة و الرضا , فعندما يحدث ذلك عليكم ان تتكلموا بحزم و احساس عميق بالمسؤالية	Query	من اليوم ذهبت افكار جديدة بدل عميق احر أن			
السابق , ان احز اني لن تثنيني عن المسير الافكار الجديدة اليوم هي العميقة بالمشاعر و Result 2 الاحز ان من الان ابدؤوا في زرع <i>افكار</i> السلام الجديدة و السعادة و الرضا , فعندما يحدث ذلك عليكم ان تتكلموا بحزم و احساس عميق بالمسؤالية	Result 1	اصبحت أفكر باسلوب جديد , لم اعد ذلك الانسان			
Result 2 والعميقة بالمشاعر و الافكار الجديدة اليوم هي العميقة بالمشاعر و الاحزان من الان ابدؤوا في زرع <i>افكار</i> السلام الجديدة و السعادة و الرضا فعندما يحدث ذلك عليكم ان تتكلموا بحزم و احساس عميق بالمسؤالية		السابق , ان احز اني لن تثنيني عن المسير			
الأحزان من الان ابدؤوا في زرع <i>افكار</i> السلام الجديدة و السعادة و الرضا فعندما يحدث ذلك عليكم ان تتكلموا بحزم و أحساس عميق بالمسؤالية	Result 2	الافكار الجديدة اليوم هي العميقة بالمشاعر و			
من الان ابدؤوا في زرع <i>افكار</i> السلام الجديدة و السعادة و الرضا فعندما يحدث ذلك عليكم ان تتكلموا بحزم و أحساس عميق بالمسؤالية		الأحزان			
السعادة و الرضا ^ت فعندما يحدث ذلك عليكم ان تتكلموا بحزم و احساس <i>عميق</i> بالمسؤالية	Result 3	من الان ابدؤوا في زرع <i>افكار</i> السلام الجديدة و			
تتكلموا بحزم و أحساس <i>عميق</i> بالمسؤالية		السعادة و الرضا فعندما يحدث ذلك عليكم ان			
		تتكلموا بحزم و احساس ع <i>ميق</i> بالمسؤالية			

Table 2 shows an example of our best system results. The query presented in Table 2 represents the baseline translation and the results are the sentences extracted from Lemur. We aim to replace inaccurate translation with most similar sentence extracted from our Arabic monolingual

corpus; extracted sentences can convey the same meaning and enjoy a correct syntactical and grammatical structure.

VI. SYSTEM FOR EXTRACTING CORRECT ARABIC TRANSLATION FROM LARGE ARABIC MONOLINGUAL CORPUS



Figure 2: System for Extracting Correct Translation

The general architecture for our correct sentence translation extraction system is show in figure 2. Our baseline which is explained is section 3 will translate from Danish into Arabic. It's trained on a small bilingual dataset of 1 M words and a DA-AR dictionary of 500 K entries. For our testing data we use a sample of 300 K words of bilingual DA-AR parallel text. We use a huge monolingual corpus of 3.0 G words. All data used in this experiment shares the same text nature of Arabic novels written in modern standard Arabic. Our baseline is trained on Moses package. We use the Arabic monolingual corpus mentioned in section 4 to build our language model, we use SRILM toolkit Stolcke (2002) for that. The baseline system receives a translation request and will produce a possible translation. We notice that the output of the translation is not accurate in some cases but that is expected due to the small training resources for our baseline. Problems with translated output like grammatical and syntactical mistakes may appear. So we search in the large monolingual Arabic corpus on a similar sentence for the translated output. We process small sentences, because GIZA++ usually ignores long sentences. In a similar approach Munteanu and Marcu (2005) used the dictionary to find the matching target sentence for the source sentence. We believe using a baseline has many advantages over this approach. For example Moses is phrase oriented decoder that studies the relationship between lexicons in the sentence to produce translation, so he can learn that "book" in "book a flight" refers to different translation than "library book". A dictionary will not be able to decide what the best translation for the word in a sentence is. Moses will produce the most probable translation and consequently will help guide our search to find the most suitable sentence in the Arabic corpus. We use Lemur tool kit for searching for candidate sentences similar to our baseline output sentence Ogilvie and Callan, (2001). We selected Lemur because it supports Arabic documents which are a great advantage unfortunately many modern IR tools doesn't have. Lemur will interact with Moses directly to produce a final translation.



Moses is trained into a small training data size and that will make its ability to produce correct sentences weak. We fix this problem by using Lemur to search for the most similar sentence to Moses output translation. Both Moses and Lemur uses the large Arabic monolingual corpus to build their own language model (LM) as described in figure 2. Moses uses it's LM to detect the most probable translation for the source sentence, while Lemur build its own LM to enhance the search process on that large monolingual corpus. Both Moses and Lemur LM's are different in structure and usage, but they use the same corpus. For Lemur to function efficiently we need to build indexes on our search space. We format our Arabic monolingual corpus according to NIST ¹ format, so that it can be recognizable from Lemur Arabic parser, table 3 gives an example of NIST style documents. We use Lemur Arabic language parsers to build Lemur indexes. After initial parsing we add a list of Arabic stopping words² to be ignored by Lemur while indexing. We use the stem indexing feature of Lemur which enhances recall results. Using this feature has shown better recall results for Lemur. We process only the top 3 scoring sentences that are returned by the IR process. We found no evidence that retrieving more than 3 top scoring sentences helped get better sentences. At the end of this step, we have for each query sentence 3 potentially matching sentences as per the IR score. The information retrieval step is the most time consuming task in the whole system. The time taken depends upon various factors like size of the index to search in, length of the query sentence etc. Ouery length also affected the speed of the sentence extraction process. We placed a limit of approximately 25 words on the queries and the indexed sentences. This choice was motivated by the fact that the word alignment toolkit Giza++ does not process longer sentences.

Table 3: NIST file structure

```
<?xml version="1.0" encoding="UTF-8"?>
<SRCSET setid="ALKARNAK_NAJEEB_MAHFOZ"
srclang="AR">
     <DOC docid="1" genre = "text" >
     <seg id="1.1">
                   اليوم يعتبر من الايام الحميلة , اعتقد أن أسامة لن
يرجع الى الاسكندرية قبل العاشرة.
     </seg>
     <seg id="1.2">
    ممكن جدا, بالمناسبة هل سنتمكن من الذهاب الى بيت سحر 👃 لقد اتم
           بي و ابلغتني انها ستكون بالبيت في المساء
     </seg>
     <seg id="1.3">
                نعم هذه فكرة جيدة
     </seg>
       </DOC>
</SRCSET>
```

VII. **CANDIDATE SENTENCE PAIR SELECTION**

When we receive the result from Lemur IR system we need to decide if the system returned sentences are parallel or not, we select sentences with the best score. We pass matched sentence pairs to the Lemur toolkit for further filters. Gale and Church (1993) alignment program is based on the assumption that longer sentences tend to be translated into longer sentences and shorter sentences tend to be translated to shorter sentences in other language.

We use the same logic for selecting our language pairs, a sentence pair is selected for further processing if the length ratio is more than 1.5, having in mind that Danish sentences are longer than their counterparts Arabic translations. We kept a relaxed factor of 1.5 as a length ratio.

Selected sentence pairs according to the previous mentioned criteria are then verified according to the WER (Levenshtein distance) and Translation Error Rate (TER). WER measures the number of operations required to transform one sentence into the other (insertions, deletions and substitutions). A zero WER means that the two sentences are identical, while lower WER means that sentence pairs are sharing most of the common words. WER can't detect a correct translation which may be different in order of the words because it work on word by word basis. We solve this shortcoming by using the TER which measures the number of edits needed to change system output into a given reference translation. TER allows words movements and thus can absorb word recording or rephrasing in translation as described by Snover et al. (2006). Generally, sentences selected based on TER filter showed better BLEU scores than their WER counter parts. So we chose TER filter as standard for our experiments with limited amounts of human translated corpus. Figure 3 shows a BLEU score comparison between WER and TER based on the size of the training data. These experiments were performed with only 1.3 M words of human-provided translations (Arabic Novels corpus).



Figure 3: System performance based on TER and WER filter

1: http://trec.nist.gov/

Published By:

2: http://arabicstopwords.sourceforge.net

EXTRACTING PARALLEL SENTENCES VIII. FROM COMPARABLE RESOURCES.

We intend to boost our baseline with more bilingual DA-AR training data .We extract parallel sentences form available comparable corpora. There is no direct bilingual DA-AR large corpus available, but there are many DA-EN and EN-AR corpora. We intend to use these resources to boost our baseline performance. We mine common English sentences between the two available groups (DA-EN, EN-AR). If we find a match then we extract the parallel Danish sentence from DA-EN group and we do the same for Arabic sentences with EN-AR group. We end up with a new parallel DA-AR sentences which can be used in to increase the size of our baseline data training. We start from the DA-EN group and we process every.



A Hybrid Model for Autonomous Danish-Arabic Statistical Machine Translation

English sentence as follow:

- We extract a window of 4-gram size as an input
- Lemur is used to decide what documents in the EN-AR group might contain this sentence (window). We build Lemur indexes only on the EN-AR side and in the same approach mentioned in section 5.1.
- We use the stopping word list provided by the IR Group of University of Glasgow¹ for Lemur index.
- We apply the Longest Common Substring problem on both source and target sentences, see Section 5.3.1.
- We extract the matching Danish and Arabic sentences as a new parallel pair. Table 5 provides an example of these steps.

IX. LONGEST COMMON SUBSTRING (LCS)

Finding a matching sentences between the two groups mentioned in section 5.3 is time consuming and a nontrivial task to tackle. We developed the algorithm shown in table 4 as an implementation for the LCS problem.



LCS-Length(Src, Dest)
L = length[Src]
H = length[Dest]
for Start = 1 to L
c[i 0] = 0
for $i = 1$ to H
c[0 i] = 0
for $i = 1$ to L
for $i = 1$ to H
$if(\mathbf{x}[i] - \mathbf{y}[i]) $
$r(x_{[1]} = y_{[1]})$
b[i,j] = b[i',j'] + 1
U[1, j] = 1
$\int \frac{1}{2} \int $
$c_{[1,1]} = c_{[1,1]}$
C[1, j] = C[1-1, j]
D[1,j]=N
}
else {
c[i,j] = c[i,j-1]
b[i,j] = W
}

The algorithm declare a two dimensional array to represent the two sentences to be matched. At the start the matrix elements are initialized with zeros as shown in table 5. Then it will search for a match on word bases, if it finds it will add 1 to the diagonal position of the cell. Finally when the Algorithm finishes all possible comparisons between words, the cell with the largest value would represent the longest sequence (sentence) out of our sentence comparisons . We extract these words (window), along with the parallel accompanied Danish and Arabic sentences. Table 4 gives an example of the algorithm major steps.

Table 5: LCS Process Example

			D	anish	i-Eng	zlisł	1					٦	
		1 D2			D3		D4						
			E	1	E2		E3		E4				
A1		E1	0		0		0		0		0		
A2		E2	0		0		0		0		0		
A3		E3	0		0		0		0		0		
A4		E4	0		0		0		0		0		
Ara	b-En	glish	0		0		0		0		0		
LCS	array	stru	cture										
				Da	aich 1	Enc	liah					_	
				Dal	nisn-	Eng		T	• • • •			-	
				INY	2	In	Indtil		vigtige		spørgs		
				Em	ero	Р	endi	i	mno	rt	Issues		
				ing	5	n	g	a	nt	- •	100400		
	العد	Ma	any										
	يد			0		0		0			0		0
	الـعـا اتـت	pe	ndi	0									
	لقه الم	ng	non	0		0		0	0		0		0
	المه مة	tai	ipor nt	0	0			1	1		0		0
ish	من	Iss	ues	0				1			Ū		
ngl	الق												
э-Е	ضا												
ral	يا			0		0	0		0		2		0
A		0		0		0			0		3		
LCS	array	valu	es aft	er ma	atchir	ng						_	_
				Dai	ush-	Eng	lish	•		1			
				Nye		In	natii		vigtige		spørgs mål		
				Em	erg	Pe	ndi	im	port		Issues		
	, 11	14	0.001	ing		ng		an	t			\parallel	_
	العد بد	IVI	any			0		0			0		0
	العا	pe	ndi	0		0		0			~		ㅓ
	لقة	ng	5	0		0		0			0	C)
	المه	In	npor										٦
_	مة	ta	nt	0		0		0		(0)
glisł	من الت	Is	sues										
Eng	الف من ا												
ab-]	ab-H		0		0			0		0	r	0	
Ar				0		0		0	0		0	ſ	$\frac{1}{2}$
Eng	lish		Pen	ding		In	iporta	ant		Issi	ues		ĺ
							r si a			Danish Indtil Vigtige Spørgsmål			
Dar	nish		Indt	11		V1	igtige	;		Spo	orgsmal		

X. RESULTS AND EVALUATIONS

Our goal was to enhance our baseline performance by replacing inaccurate Arabic translation with more accurate one using large Arabic monolingual corpus. In this section we report the results of using this approach with our baseline. We conduct many experiments to measure the effect of using this process to translation quality and correctness. Table 6 shows a comparison of two baselines one developed with translation. correction approach and the other without it.



^{1:}http://ir.dcs.gla.ac.uk/resources/linguisticutils/stopwords

The sentence correction approach best BLEU score was 20.63 which is +2.51 points better than the other baseline that's best BLEU score was 18.12. As expected both systems performs better when the training data increases. The correction of translation approach has shown that it is possible to produce better translation system. In the second experiment we train our baseline with the extracted text from our parallel sentences mining step discussed in section 5.3. Table 7 shows the BLEU scores of this experiment which was time consuming, but time efficiency was not a concern for us compared to the extracted sentences output. We managed to extract 500 K words from all the available resources mentioned in section 4.

	·	
Size	BLEU Baseline	BLEU Baseline + Translation Correction
200 K	12.17	18.23
400 K	14.23	18.30
600 K	16.36	18.22
800 K	17.52	20.42
1.3 M	18.12	20.63

Table 6: System BLEU scores

Our baseline system BLEU scores always increased after adding this new parallel training data. We observed that our system best score was 21.13. The addition of the dictionary data entries to the baseline performance had more positive effect (+15%) than the extracted data from comparable resources (+2%). We think this is because of the different nature of the extracted data (legal text) from the training data.

Table 7: Effect of extracted data to our system performance.

Data	Size	BLEU
Novels	0.8 M	17.41
Novels + Dictionary	1.3 M	20.63
Novels + Dictionary+ Extracted data	1.8 M	21.13

CONCLUSION AND FUTURE WORK XI.

Our motivation for this approach was to be able to improve our DA-AR SMT performance by inspecting two factors; first one is the ability to enhance the translation output by replacing the baseline output Arabic translation with a similar sentence that we retrieve from large Arabic monolingual corpus. The second factor we tackled was enhancing the baseline with parallel DA-AR text that we mine from parallel DA-EN and EN-AR resources. The lack of parallel DA-AR resources has encouraged us to think of using existing language resources available to create DA-AR parallel resources. Our experiments results indicate the validity of building a baseline with translation correction approach. Our approach mainly uses incomplete knowledge (translated sentence) to produce a complete Knowledge. Large bilingual resources are not required to produce a reasonable translation. A major focus is on the large Arabic monolingual corpus used to build Moses language model to extract correct sentences. Our system produces translations that are more "human oriented" rather than "literally oriented". Translations would carry the same meaning as source sentence but maybe with difference in sentence

structure or words order but it would still hold the meaning. We are interested in using the baseline output in training other similar baseline. We plan to use more intelligent approaches in detecting similar sentence in the Arabic monolingual corpus. We will apply a more syntactical Analysis to the baseline output before searching for similar sentences like removing connected articles from the Arabic sentence and search for the word and its concepts or synonyms.

REFERENCES

- 1. Stolcke. SRILM- an extensible languagemodeling toolkit. 2002. In Proc. Int. Conf. on Speech and Language Processing (ICSLP), volume 2, pages 901-904, Denver
- 2. Almut Silja Hildebrand, Matthias Eck, Stephan Vogel, and Alex Waibel 2005. Adaptation of the Translation Model for Statistical Machine Translation based on Information Retrieval. Proceedings of EAMT 2005: 133-142.
- Arne Mauser, Richard Zens, Evgeny Matusov, Sasa Hasan, Hermann 3. Nev 2006. The RWTH Statistical Machine Translation System for the IWSLT 2006 Evaluation. Proceedings of International Workshop on Spoken Language Translation.:103-110
- Bing Zhao, Matthias Eck, Stephan Vogel 2004. Language Model 4 Adaptation for Statistical Machine Translation with structured query models. COLING-2004
- Christopher C. Yang and Kar Wing Li. 2003. Automatic construction 5. of English/Chinese parallel corpora. J. Am. Soc. Inf. Sci. Technol., 54(8):730-742
- 6. Dragos StefanMunteanu and DanielMarcu. 2005. Improvingmachine translation performance by exploiting non-parallel corpora. Computational Linguistics, 31(4):477-504.
- Douglas W. Oard. 1997. Alternative approaches for cross-language 7 text retrieval. In In AAAI Symposium on Cross-Language Text and Speech Retrieval. American Association for Artificial Intelligence.
- 8. Franz Josef Och and Hermann Ney. 2002. Discriminative training and maximum entropy models for statistical machine translation. In ACL, pages 295-302.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison 9. of various statistical alignment models. Computational Linguistics, 29(1):19-51.
- 10. Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In ACL.
- 11. Masao Utiyama and Hitoshi Isahara. 2003. Reliable measures for aligning Japanese-English news articles and sentences. In Erhard Hinrichs and Dan Roth, editors, ACL, pages 72-79.
- Matthias Eck, Stephan Vogel, and Alex Waibel 2004. Language 12. Model Adaptation for Statistical Machine Translation Based on Information Retrieval. Proceedings of Fourth International Conference on Language Resources and Evaluation: 327-330
- Matthias Eck, Stephan Vogel, Alex Waibel 2005. Low cost portability 13 for statistical machine translation based on n-gram coverage. MT Summit X: 227-234
- 14 Pascale Fung and Percy Cheung. 2004. Mining very nonparallel corpora: Parallel sentence and lexicon extraction via bootstrapping and em. In Dekang Lin and Dekai Wu, editors, EMNLP, pages 57-63, Barcelona, Spain, July.
- Philipp Koehn et al. 2007. Moses: Open source toolkit for statistical 15 machine translation. In ACL, demonstrationsession.
- Paul Ogilvie and Jamie Callan. 2001. Experiments using the Lemur 16. toolkit. In In Proceedings of the Tenth Text Retrieval Conference (TREC-10), pages 103-108.
- Philip Resnik and Noah A. Smith Y. 2003. The web as a parallel 17. corpus. Computational Linguistics, 29:349-380.
- 18 Sadaf Abdul-Rauf , Holger Schwenk, On the use of comparable corpora to improve SMT performance, Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Greek 2009,16-23,
- 19. William A. Gale and Kenneth W. Church. 1993. A program for aligning sentences in bilingual corpora. Computational Linguistics, 19(1):75-102.
- 20. Ying Zhang, Almut Silja Hildebrand, Stephan Vogel 2006. Distributed Language Modeling for N-best List Re-ranking. EMNLP-2006: 216-223



Published By: