Adapting a Medical Speech to Speech Translation System (MedSLT) to Arabic

Pierrette Bouillon
University of Geneva, TIM/ISSCO, ETI
40, Bd. Du Pont d'Arve
CH-1211 Geneva 4, Switzerland
Pierrette.Bouillon@issco.unige.ch

Sonia Halimi
University of Geneva, TIM/ISSCO, ETI
40, Bd. Du Pont d'Arve
CH-1211 Geneva 4, Switzerland
Sonia.Halimi@eti.unige.ch

Manny Rayner
Powerset Inc
475 Brannan Str.
San Francisco
CA 94107, USA
manny@powerset.com

Beth Ann Hockey
Mail Stop 19-26, UCSC UARC, NASA
Ames Research Center, Moffett Field,
CA 94035-1000
bahockey@ucsc.edu

Abstract

We describe the adaptation for Arabic of the grammar-based MedSLT medical speech system. The system supports simple medical diagnosis questions about headaches using vocabulary of 322 words. We show that the MedSLT architecture based on motivated general grammars produces very good results, with a limited effort. Based on the grammars for other languages covered by the system, it is in fact very easy to develop an Arabic grammar and to specialize it efficiently for the different system tasks. In this paper, we focus on generation.

1 Introduction

MedSLT is a medical speech translation system. It allows a doctor to ask diagnosis questions in medical subdomains, such as headaches, abdominal pain, etc, covering a wide range of questions that doctors generally ask their patients. The grammar-based architecture, built using specialization from reusable general grammars, is designed to allow a rapid development of different domains and languages. Presently, it supports English, French, Japanese, Spanish and Catalan. This article focuses on the system development for Arabic.

In general, translation in this context raises two specific questions: 1) how to achieve recognition quality that is good enough for translation, and 2) how to get translations to be as idiomatic as possible so they can be understood by the patient. For close languages and domains where accuracy is not very important (e.g. information requests), it may be possible to combine a statistical recognizer with a commercial translation system as it is often done in commercial tools such as *SpokenTranslation* (Seligman and Dillinger, 2006). However, for this specific application in a multilingual context, this solution is not applicable at all: even if perfect recognition were possible (which is far from being the case), current commercial tools for translating to Arabic do not guarantee good quality. The domain dealt with here contains, in fact, many structures specific to this type of oral dialogue that can not be handled by these systems. For example, all the doctor's interactions with the MedSLT system consist of questions whose structures differ from one language to another, with each language having its own constraints. Consequently, two types of errors occur in Arabic translation systems. Either they do not recognize the interrogative structure as in example (1), or they produce ungrammatical sentences by copying the original structure as in example (2):

Proceedings of the 5th Workshop on Important Unresolved Matters, pages 41–48,
Prague, Czech Republic, June 2007. ©2007 Association for Computational Linguistics

(1) was the pain severe?
كان الألم الشديد (***Google***)
(*kana al alam chadid*)
'have-past-3 the pain severe'

is the pain aggravated by exertion?
الألم يفاقم بجهد (***Systran***)
(*al alam yufaqim bi juhd*)
'the pain escalate-3 with effort'

(2) is the headache aggravated by bright light?
لا يضيء بشكل ساطع صداعين سبب (***Cimos***)
(*la yudhi' bi chakl sathi' suda'ayn sabab*)
'not light in manner bright-3 headache-plur cause'

is the headache aggravated by bright light?
يتم ساطعة خفيفة سبب صداعات (***Systran***)
(*yatim sathia khafifa sabab suda'at*)
'finish-3 bright-fem not-heavy-fem cause headache-plur'

are your headaches accompanied by nausea?
إن ترافق صداعاتك بواسطة غثيان
(***1-800-translate***)
(*in turafiq suda'atik bi wasithat rhathayan*)
'if you-accompany your headache-plur using nausea'

Ellipsis is another problem for MT systems. Many elliptical structures cannot be translated without context. In example 3, the context is needed to guarantee adjective agreement.

(3) Doctor: is the pain severe?
Trad: هل الألم شديد؟ (MedSLT)
(*hal al alam chadid*)
'Q the pain severe'

Doctor: moderate?
Trad: محتمل؟ محتملا؟ محتملة؟
(*muhtamala, muhtamalan, muhtamal*)
'moderate_fem_attributive_adj, moderate_vocalized-predicative_adj, moderate_attributive_adj'.

It is also essential for rules of translation to be applied consistently. For instance, in MedSLT, *onset* is translated by the verb ظهر (*dhahara*). In this context, the adjective *sudden* has to be translated by an adverb فجأة (*fajatan*) (example 4). This implies that the translation of the ellipsis in the second utterance needs to change syntactic category too. We can wonder to what extent the word-for-word translation of the elliptical sentence in (4) can be understood by the patient.

4) Doctor: was the onset of headaches sudden?
Trad: هل ظهر الصداع فجأة؟ (MedSLT)
(*hal dhahara al sudaa fajatan?*)
(Q appear-past-3 the headache suddenly?)

Doctor: acute?
Trad: مفاجئ؟
(*mufaji?*)
(acute)

In addition to that ellipsis can not always be translated by the same type of ellipsis. Arabic grammar (Amin, 1982) allows the use of elliptical structures in cases where there is a semantic link (قرينة) referring to the omitted part of the sentence otherwise the elliptical construction is ambiguous. In example (3), the use of an adjective alone presents an ambiguity introducing, therefore, a difficulty in comprehension which can be problematic. Thus, it is necessary to resort to a more sophisticated approach. We will describe, in the following part, the architecture on which MedSLT is based. Then, we will show how it has been adapted to Arabic.

2 The Architecture

MedSLT is a grammar-based medical speech translation system which uses the commercial *Nuance* speech recognition platform. It has two main features (Bouillon et *al.*, 2005). First, all the language models (for recognition, analysis, generation) are produced from linguistically motivated, general unification grammars using the *Regulus* platform (Rayner, et *al.*, 2006). First, domain specific unification grammars are created from the general grammar for the different domains of medical diagnosis through a trainable corpus-based automatic grammar specialization process. They are, next, compiled into Context Free Grammars (CFGs) in a format suitable for use with the *Nuance* speech

recognition platform, and into a form needed for a variant of Semantic Head-driven generation (Shieber et *al.,* 1990). Therefore, the different grammars needed by the system under this approach are easy to build and maintain.

This leads us to the second feature. Because grammar-based speech recognition only produces competitive results for the sentences covered by the grammar, the user will need to learn the coverage of the system. In order to assist in this, a help system is included in the system (Starlander et *al.*, 2005 and Chatzichrisafis et *al.*, 2006). The help system suggests, after each user utterance, similar utterances covered by the grammar which can be taken as a model. In order to derive the help sentences, the system performs, in parallel, a statistical recognition of the input speech. It then compares the recognition result using an N-gram based metric, against a set of known correct in-coverage questions to extract the most similar ones. It is in that way that we introduce some of the robustness of the statistical systems in the controlled application.

Once the sentence recognized, the translation is interlingua-based. *Regulus* allows different types of source representations (Rayner, et *al.*, 2006), but we have chosen to use the simplest one in order to facilitate the translation process. It is a flat semantic structure built up by concatenation of word meanings. For instance, هل يشتد الصداع عند القلق؟ (*hal yachtaddou al soudaa inda al qalaq?* 'Q aggravate-3 the headache in the stress') would be represented as follows:

```
[[cause,qalaq],[event,yachtaddou],
[prep_cause,inda],[symptom,soudaa],
[tense,present],[utterance_type,ynq],
[voice,active]]
```

The same formalism is used for the interlingua which is a standardized version of the most explicit source English representation. For example, the interlingua representation of the previous sentence corresponds to the following structure that can be paraphrased as follows: « does the pain become worse when you experience anxiety?»:

```
[[sc,when],[clause,
       [[pronoun,you],
       [secondary_symptom,anxiety],
       [state,experience],
       [tense,present],
       [utterance_type,dcl],
       [voice,active]]],
[event,become_worse],[symtom,headache],
[tense,present],[utterance_type,ynq],
[voice,active]].
```

Under this approach the translation process only involves mapping simple structures. This facilitates the process of translation and the resolution of divergences. This process goes through five stages: 1) source language analysis in order to extract source representation; 2) ellipsis resolution if necessary; 3) mapping the source structure into the interlingua; 4) mapping the interlingua into the target structure and 5) generation of the target language in accordance with its own grammar.

We will show next the adequacy of this architecture for translation in Arabic. On the basis of the grammars already implemented for some languages covered by the system (French, English, Spanish, Catalan), it is, in fact, easy to develop a general Arabic grammar that meets the constraints of the MedSLT project and to specialize it for the purposes of speech recognition and generation. This method produces very good results when compared to commercial systems.

3 General MedSLT grammar for Arabic

Writing unification grammars for speech presents two requirements. Since it has to be transformed into context–free grammar (CFG) for recognition, features must have a finite number of values, as limited as possible. In practice, this means that attributes can not take complex values and the lexicalist approach used in LFG or HPSG cannot be applied here. For example, subcategorization is not represented with general rule schemata as in HPSG. Therefore, syntagmatic rules must be multiplied for each type of verb (transitive, intransitive, etc.). Even if this first constraint results in a less elegant and more repetitive grammar, it is not a limitation to the development of grammars with the complexity required for such applications.

The grammar is used to constrain the recognition process, so it needs to include all information that could improve recognition. For instance, evaluation has shown that the quality of recognition decreases considerably when selection restrictions are omitted (Rayner, et *al.*, 2006). Thus, in practice, this means that all *Regulus* general grammars include many features for managing this

type of constraint. For example, nouns are semantically typed; verbal entries contain features to determine the type of complements according to their subcategorization, etc. These types are difficult to define coherently for the general vocabulary but are not problematic when the domain is very controlled and the vocabulary very limited. In addition, they do not have any effect on the whole structure of the general grammar since they come from specialized lexica of various domains.

As with all Regulus grammars, the Arabic grammar and lexicon are bound by these two restrictions. At the present time, they cover only questions in relation to headaches. The vocabulary contains 322 different forms. Nouns are semantically typed and verbs specify the type of complements. For instance, the entry أجريت (*ajrayta,* 'carry out') indicates that the verb selects a subject which is an agent (**subj_np_type=agent**) and an object of **thera** type (therapeutic) (**obj_np_type=thera**):

```
v:[sem=[[state,tajri],[tense,passé]],
   subcat=trans, agr=2/\sing/\masc,
   vform=finite, subj_np_type=agent,
   obj_np_type=thera] -->
        @a('أجريت', ajrayta).
```

It is interesting to note that features and values are the same in Arabic as in other languages except for some differences such as the **agr**(eement) feature which can take a "dual" value, *inter alia*. To avoid the multiplication of entries, particles such as ال (*al*), بـ (*bi*), ك (*ka*), were separated from words to which they are normally attached. For recognition, this does not seem to pose a problem. For generation, they are joined to their heads according to specific orthography rules after the generation of sentences. Since the word is synthesized, it appears only in its non-vocalic form.

The grammar contains 38 rules that describe yes-no questions introduced by هل (*hal*), for example : هل يمتد الألم إلى الكتفين؟ (*hal yamtad al alam ila al katifayn,* 'Q irradiate-3 the pain to the shoulders') and some wh-questions, for example : متى يظهر الألم؟ (*mata yadhar al alam,* 'when appear-3 the pain'). The grammar structure is, in the end, quite close to romance languages. As it can happen in Spanish or Catalan, the subject of yes-no questions in Arabic comes conventionally after the verb (*hal yamtad* [*al alam*]_sujet [الألم]) if not elided when it is agentive (*hal* [*tahus*] *bi al alam* [تحس], 'Q you-feel with the pain'). Thus, we can use similar rules applied to *Prodrop* and *inversion* in these languages. Inversion is not dealt with as a type of movement otherwise it would have obliged us to multiply the number of features in the grammar. Instead, we use the constituent **vbar**, which is also convenient for Arabic. We consider a yes-no question (**yn_question**) to be made up of a particle, which is هل (*hal*), and a sentence (**s**) where the subject is either elided (**inv=prodrop)**, or comes after the verb (**inv=inverted**), namely:

```
yn_question:[sem=...] -->
       @a('هل', hal),
       optional_adverb:[...],
       s:[...inv=inverted\/prodrop].
```

The **s** is rewritten in a **vp** which is itself constituted of a **vbar** and its complements according to the type of the verb (transitive, intransitive, etc.) as is a standard grammar structure:

```
s:[sem=] -->
       vp:[inv=INV, …].

vp:[sem=…, inv=INV, …] -->
       vbar:[subcat=trans, inv=INV, ….],
       optional_adverb:[….],
       np:[….],
       optional_adverb:[…],
       optional_pp:[….],
       optional_adverb:[….].
```

The **vbar** is itself composed of a single verb (if the subject is elided; in this case it has an **inv=prodrop** feature), or a verb followed by a subject (in such instance, it has a **inv=inverted** feature) as in rules shown above. We note that the elision will only be possible here if the verb takes a subject of **agent** type:

```
vbar:[sem=…,inv=prodrop] -->
       optional_v:[agr=2/\masc/\sing,
       subj_np_type=agent].


vbar:[sem=..,inv=inverted] -->
       optional_v:[],
       np:[].
```

The treatment of *wh*-questions is more conventional in all languages because it is not possible to handle them without simulating movement. We consider that the interrogative pronoun moved

from its initial place (PP, etc.), which becomes empty, to an initial position (*ayna_i tahus bi al alam [i]*, أين تحس بالألم, 'where you-feel with the pain'). To deal with the movement, we use the standard mechanism of gap threading, introduced by Pereira (1981). The link between the empty constituent [i] and the constituent which has been moved (*ayna_i* in our example) is possible using two attributes which are **gapsin** and **gapsout**, included in all categories related to the movement. For example, in the following rule, such attributes indicate that the interrogative element (**wh_pp**) is only possible if the sentence (**s**) contains an empty pp (indicated by the attribute **gapsin=pp_gap**):

```
wh_question:[sem=...] -->
       wh_pp:[sem=..],
       s:[…, gapsin=pp_gap, gapsout=B].
```

In comparison with the rest of languages previously processed by the system, the Arabic grammar does not have a lot of special cases. One rule specifies that some verb such as "to be" (كان (*kana*), يكون (*yakun*), with the feature subcat=pred(icatif)) can be optional – they can be rewritten in an empty constituent indicated as []:

```
optional_v:
       [sem=[[state,be],[tense,present]],
       subcat=pred] --> [].
```

Rules for numbers are also very complex in order to represent the dual form in addition to the position of numbers which can change depending on whether the number is singular: one, for example : more than one day أكثر من يوم واحد (*akthar min yawm wahid,* 'more than day one') and, the third day اليوم الثالث (*al yawm al thalith,* 'the day the third'), or plural, for example : more than 3 days, أكثر من ثلاثة أيام (*akthar min thalathat ayam,* 'more than three days').

4 Grammar specialization

One of the most important advantages of the approach adopted here is that the general grammar can be specialized for use in different tasks and domains, to obtain a useful grammar in each case. In the case of Arabic, it is possible to perform generation and recognition directly using the general grammar described above, since it is not yet very elaborate. The general grammar is however already large enough to cause serious efficiency problems, When compiled for generation, the general grammar overgenerates, as the target structures are flat and underspecified (they do not include, for example, information on numbers or determiners, cf. examples above). It would be possible to insert preference rules to force the intended structure, but this solution is extremely unattractive from a software engineering point of view. When compiling the grammar for recognition, the situation is even worse. All our experiments on other languages show that recognizers compiled from general grammars either perform very poorly (Bouillon et al 2007), or fail to recognize at all (Rayner et al 2006, section 11.7). As in previous work, we have attacked these problems by creating specialized versions of the general Arabic grammar.

In our approach to grammar specialization, domain-specific unification grammars are derived from general ones using the Explanation Based Learning (EBL) technique (Rayner, et *al.*, 2006). This corpus-based specialization process is parameterized by a training corpus and a set of operationality criteria. The training corpus, which can be relatively small, consists of examples of utterances that should be covered by the domain grammar. (For Arabic, the current training corpus is about 450 sentences). The sentences of the corpus are parsed using the general grammar, then those parses are partitioned into phrases based on the operationality criteria. Each phrase defined by the operationality criteria is flattened, producing rules of a phrasal grammar for the application domain. The resulting domain-specific grammar has a subset of the coverage of the general grammar and reduced structural ambiguity. In a generation grammar, over-generation is virtually eliminated; specialized recognition grammars typically have greatly superior recognition due to the reduction in search space that they provide. In the case of the Arabic grammar described here, the training corpus is a set of Arabic sentences based on the English reference corpus for the headaches domain. The operationality criteria are a slightly modified version of those used for the Romance grammars discussed in Bouillon et al., 2007.

In previous work, we have described at length the structural relationships between general grammars, and specialized grammars for recognition and generation; here, we will briefly summarize

the main points and show a simple example of how they apply to our Arabic grammar. Figures (1) to (3) present parse trees for the sentence هل الألم دائم؟ (*hal al alam daym* 'Q the pain permanent'):

```
.MAIN
   utterance
      yn_question
      /  lex(hal)
      |  optional_adverb null
      |  s
      |     vp
      |     /  vbar
      |     |  /  optional_v null
      |     |  |  np
      |     |  |  /  spec lex(al)
      |     |  |  |  nbar
      |     |  \  \     noun lex(alam)
      |     |  adj lex(daym)
      |     |  optional_adverb null
      |     |  optional_pp null
      \     \  optional_adverb null
```

Figure (1): Parse tree for *'hal al alam daym'* with the general grammar

```
.MAIN
   utterance
   /  lex(hal)
   |  vp
   |  /  vbar
   |  |     np
   |  |     /  spec lex(al)
   |  |     \  noun lex(alam)
   |  |  adj lex(daym)
   \  \  optional_pp null
```

Figure (2): Parse tree for *'hal al alam daym'* with the specialized recognition grammar

```
.MAIN
   utterance
   /  lex(hal)
   |  vp
   |  /  vbar
   |  |     np lex(al) lex(alam)
   |  |  adj lex(daym)
   \  \  optional_pp null
```

Figure (3): Parse tree for *'hal al alam daym'* with the specialized generation grammar

It is immediately apparent that (1), the parse tree for the general grammar, is structurally much more complex than (2) and (3), the trees for the specialized grammars. In particular, (1) has several nodes filled by optional modifiers of various kinds, all of which are here null; if this grammar is compiled into a recognizer, all these nodes result in extra paths in the search space, with a corresponding loss of efficiency. Both the specialized grammars flatten out the modifier structure, for example using learning a set of vp rules which instantiate only those combinations of modifiers that have actually been seen in the training corpus.

The difference between the specialized recognition grammar (2) and the specialized generation grammar (3) is more subtle. The first thing to consider is that the recognition version needs to contain all the rules required for recognition and analysis of multiple syntactic variants of the diagnosis questions, while the generation one only has to contain sufficient rules to generate one variant (ideally, the most correct and idiomatic one) for each question. An important consequence of this general principle relates to the treatment of NPs. The general grammar includes a rule that forms an NP in a conventional manner from a specifier (*iddat, koul* and *al*, which has been separated from the noun), potentially optional, and a noun. This rule permits a compositional analysis of all the grammatical combinations of nouns and articles, which is also appropriate for the recognition grammar. For generation, however, the system learns generally complete (lexicalized) NPs, in order to attach the appropriate article for each noun on the basis of the corpus (there is an exception for NPs containing a number because it is obviously undesirable to include in the corpus one example of every number/noun combination). Contrasting (2) and (3), we see that in (2), the phrase الألم (*al alam*, 'the pain') is treated compositionally; in (3), it is a lexicalized phrase produced by the rule

```
np --> الألم          (al alam)
```

Our previous experience with French, English and Spanish has shown that this method is a good solution for specialized and limited domains like the one under study. Articles are difficult to recognize (they are usually short unstressed words) and to translate, but the right combinations can easily be learnt according to the context and subdomain. In the next section, we show that the specialization method yields good results in practice when applied to Arabic.

5 Evaluation

Our initial evaluation only tests the specialized Arabic generation grammar. We used an English corpus of 522 diagnostic questions gathered with MedSLT, which has previously been used to compare linguistic and statistical recognition (Rayner et *al.*, 2004). Translation were judged by four Arabic translators from the Geneva Translation School on the following three-point scale:

- **Good** : translation respects completely the meaning and the grammatical form;
- **OK** : translation is not completely idiomatic but understandable;
- **Bad** : translation does not keep the meaning, is non understandable or it is agrammatical.

The results are as follows:

Evaluation	T1	T2	T3	T4
Good	365 (69.9%)	318 (60%)	323 (61%)	281 (53%)
Ok	16 (3.1%)	63 (12%)	56 (10%)	86 (16%)
Bad	3 (0.6%)	3 (0.6%)	5 (0.9%)	17 (3%)
Not analyzed sentences	114 (21.8%)			
Not translated sentences	21 (4%)			
Not generated sentences	3 (0.6%)			
Total	**522 (100.0%)**			

We clearly can see that translations are good (**Good** or **Ok**) if the sentences are well recognized/analyzed in English, which is very important for our application (381/408 for **T1** (93%), 381/408 for **T2** (93%), 379/408 for **T3** (92%), 367/408 for **T4**, (89.9%)). **Not analyzed sentences** (21.8%) are those which are not covered by the English grammar but had to be reformulated in an existent structure with the help system (see above; Chatzichrisafis, et *al.*, 2006).

Three sentences only (0.6%) failed at the level of generation (**Not generated sentences),** which shows that the specialized generation grammar is robust enough for the domain. These sentences have now to be added in the corpus to be generated correctly. In other languages, we have indeed noticed that this kind of error disappears after one or two cycles of evaluation on new data. **Not translated sentences** (4%) are mostly caused by specialized medical terms describing pain (*pounding, throbbing*, etc.) that we did not introduce yet because they need to be validated by Arabic medical specialists. Here are some examples of **Good** translations:

does chocolate cause your headaches
هل يظهر الصداع عندما تأكل الشوكولا؟
(*hal yadhharou al soudaa indama takoul al chocolat*)
(Q appear-3 the headache when you-eat the chocolate)

do headaches usually occur in the morning
هل كثيرا ما تحس بالصداع في الصباح؟
(*hal kathiran ma tahus bi al soudaa fi al sabah*)
(Q often ma-you-feel-bi the headache in the morning)

is the headache in the front of your head
هل تحس بالصداع في الجبهة؟
(*hal tahus bi al soudaa fi al jabha*)
(Q you-feel-bi the headache in the front)

does stress cause your headaches
هل يظهر الصداع عند الإرهاق؟
(*hal yadhharou al soudaa inda al irhaq*)
(Q appear-3 the headache in the stress)

is it a stabbing pain
هل الألم مثل طعنة سكين؟
(*hal al alam mithl taanat sikin*)
(Q the pain like stabbing knive)

In order to compare our results with commercial MT systems output, we submitted the first 124 well analyzed sentences to *Systran.* Among these translations, 98 were judged as **Bad**, 6 as **Good** and 20 as **Ok**. What the translator has considered as bad are the translations that are not in the interrogative form and neither grammatical nor idiomatic. Consequently they are not understandable. Here are the first ten translations we have obtained:

Original sentence (English)	Translation (Arabic)	Evaluation
is the pain relieved by stress removal?	يكون الألم خفّفت بإجهاد إزالة ؟ *(yakun al alam khafafat bi ijhad izalat)* 'be-3 the pain relieve-	Bad

	past-fem with stress removal'	
does the pain extend to your neck?	الألم يمدّد إلى عنقك ؟ *(al alam yumadid ila ounoukika)* 'the pain make-longer-3 to neck-yours'	Ok
is the pain severe?	يكون الألم قاسية ؟ *(yakun al alam qassiya)* be-3 the pain harsh-fem'	Ok
is caused by bright light?	سبّبت بضوء ساطعة ؟ *(sababat bi dhaw sathia)* 'cause-she with light bright-she'	Bad
is the pain made better by coffee?	يكون الألم يجعل جيّدة بقهوة ؟ *(yakun al alam yajaal jayida bi qahwa)* 'be-3 the pain make good-fem-3 with coffee'	Bad
does it sometimes last more than two hours?	هو أحيانا يدوم أكثر من اثنان ساعات ؟ *(howa ahyan yadum akthan min ithnan saat)* 'he sometimes do-last-3 more than two hours'	Bad
do you have headaches in the morning?	أنت تتلقى صداعات في الصباح ؟ *(anta tatalaqa sudaaat fi al sabah)* 'you do-receive headache-plur in the morning'	Bad
how long do your headaches last?	[هوو لونغ] يتمّ صداعاتك يدوم ؟ *([how long] yatim sudaatik yadum)* 'how long finish-3 headache-plur- yours long'	Bad
thirteen minutes to a few hours?	ثلاثة عشر دقائق إلى [ا فو] ساعات ؟ *(thalatat achar daqaiq ila [fu] saat)* 'thirteen minutes to saat'	Bad
how long does the headache last?	[هوو لونغ] الصداع يدوم ؟ *([how long] al sudaa yadum)* '[how long] the headache last-3'	Bad

6 Conclusion

At the present time, it would have been difficult to use a commercial machine translation system for Arabic in the context of our application where accuracy is very important. One possibility is thus to use a more linguistic approach that takes advantage of the subdomain constraints. This approach is usually very costly. However we have shown in this paper that the MedSLT architecture based on motivated general grammars produces very good results, with a limited effort. The general grammar can be developed very easily on the basis of other languages. The experiments described here show good results for the Arabic generation grammar. Our initial anecdotal results with the Arabic recognizer are promising, and we hope to begin evaluation of this component in the near future.

References

P. Bouillon, F. Ehsani, R. Frederking and M. Rayner (Eds.) 2006. *Medical Speech Translation.* Proceedings of the Workshop. HLT/NAACL-06, New York, NY, USA.

P. Bouillon, M. Rayner, B. Novellas, M. Starlander, M. Santaholma, Y. Nakao and N. Chatzichrisafis. 2007. Une grammaire partagée multi-tâche pour le traitement de la parole : application aux langues romanes. *TAL.*

P. Bouillon, M. Rayner, N. Chatzichrisafis, B.A. Hockey, B.A., M. Santaholma, M. Starlander, H. Isahara, K. Kanzaki, and Y. Nakao. 2005. A generic Multi-Lingual Open Source Platform for Limited-Domain Medical Speech Translation. *Proc. 10th EAMT*. Budapest, Hungary.

N. Chatzichrisafis, P. Bouillon, M. Rayner, M. Santaholma, M. Starlander, B. A. Hockey. 2006. Evaluating Task Performance for a Unidirectional Controlled Language Medical Speech Translation System. In (Bouillon et al, 2006)

M. Rayner, B. A. Hockey and P. Bouillon. 2006. *Putting Linguistics into Speech Recognition: The Regulus Grammar Compiler.* Stanford University Center for the Study of language and information, Stanford, California.

S. Shieber and G. van Noord and F.C.N. Pereira and R.C. Moore. 1990. Semantic-Head-Driven Generation. *Computational Linguistics*, 16(1).

M. Seligman and M. Dillinger. 2006. Usability Issues in an Interactive Speech-to-Speech Translation System for Healthcare. In (Bouillon et al, 2006).

M. Starlander, P. Bouillon, N. Chatzichrisafis, M. Santaholma, M. Rayner, B.A. Hockey, H. Isahara, K. Kanzaki, Y. Nakao. 2005. Practising Controlled Language through a Help System integrated into the Medical Speech Translation System (MedSLT). *Proceedings of the MT Summit X*, Phuket, Thailand.

B. Amin. 1982. *Al-Balagha Al-Arabia. 'ilm Al-Ma'ani. Dar Al-'ilm Li-Almalayeen.* Beirut, Lebanon.