

# Rescoring N-Best Hypotheses for Arabic Speech Recognition: A Syntax-Mining Approach

**Dia AbuZeina**

Palestine Polytechnic University, P.O.Box  
198, Hebron, Palestine  
abuzeina@ppu.edu

**Husni Al-Muhtaseb**

King Fahd University of Petroleum and  
Minerals, P.O.Box 5066, Saudi Arabia  
muhtaseb@kfupm.edu.sa

**Moustafa Elshafei**

King Fahd University of Petroleum and  
Minerals, P.O.Box 405, Saudi Arabia  
shafei@mit.edu

**Wasfi Al-Khatib**

King Fahd University of Petroleum and  
Minerals, P.O.Box 5066, Saudi Arabia  
wasfi@kfupm.edu.sa

## Abstract

Improving speech recognition accuracy through linguistic knowledge is a major research area in automatic speech recognition systems. In this paper, we present a syntax-mining approach to rescore N-Best hypotheses for Arabic speech recognition systems. The method depends on a machine learning tool (WEKA-3-6-5) to extract the N-Best syntactic rules of the Baseline tagged transcription corpus which was tagged using Stanford Arabic tagger. The proposed method was tested using the Baseline system that contains a pronunciation dictionary of 17,236 vocabularies (28,682 words and variants) from 7.57 hours pronunciation corpus of modern standard Arabic (MSA) broadcast news. Using Carnegie Mellon University (CMU) PocketSphinx speech recognition engine, the Baseline system achieved a Word Error Rate (WER) of 16.04 % on a test set of 400 utterances ( about 0.57 hours) containing 3585 diacritized words. Even though there were enhancements in some tested files, we found that this method does not lead to significant enhancement (for Arabic). Based on this

research work, we conclude this paper by introducing a new design for language models to account for longer-distance constrains, instead of a few proceeding words.

## 1 Introduction

Improving speech recognition accuracy through linguistic knowledge is a major research area in speech recognition (ASR) systems. Three knowledge sources are usually presented in an ASR: acoustic models, a dictionary, and a language model as shown in Figure 1. These independent knowledge sources, also called ASR database, are subject to adapt to fulfill some natural variations that occur in speech signals. Despite that most of the adaptation occurs in the dictionary, a high integration among the ASR components is required to achieve better performance.

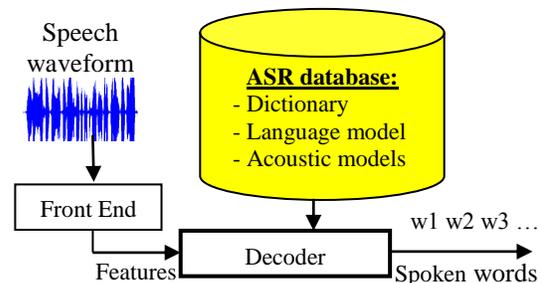


Figure 1. An ASR components

In addition to the pronunciation variation problem, the syntactic structure of the output sentence might be wrong. This problem appears in the form of taking different orders of words and phrases, out of the Arabic correct syntactic structure. Jurafsky and Martin (2009) demonstrated a reason for such phenomenon. They illustrated that variants included in the dictionary may lead to sub-optimal results which can be enhanced using N-Best hypotheses rescoring process. Jurafsky and Martin showed that the Viterbi algorithm is an approximation algorithm. This means that the Viterbi algorithm is biased against words with many pronunciations. The reason for this is that the probabilities' mass is split up among different pronunciations. In Figure 2, the system output, intuitively, is the first hypothesis while the correct output is the second one, which is highlighted. The sentences in Figure 2 are called N-Best hypotheses (also called N-Best list). In this case N is equal 5.

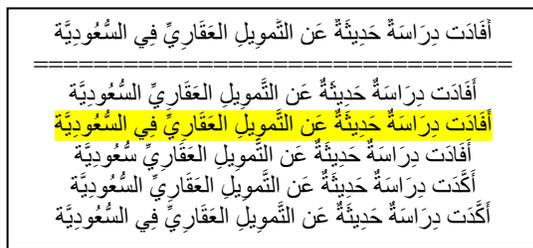


Figure 2. An example of 5-Best hypotheses

To model this problem, the tags of the words will be used as a criterion for rescoring and sorting the N-Best list. We used “language syntax rules” to indicate for the most frequently tags relationships used in the language. The rescored hypotheses are then sorted according to a new weighted scores (acoustic score and syntactic score) to pick the top score hypothesis. Figure 3 shows the idea behind the proposed rescoring model.

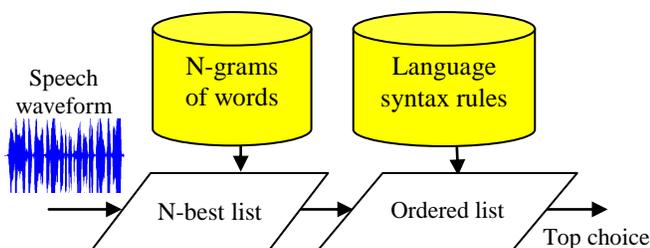


Figure 3. Illustration of rescoring N-Best list

In this work, we utilized the large vocabulary, speaker independent natural Arabic Speech Recognition system developed at King Fahd University of Petroleum and Minerals (KFUPM), based on Carnegie Mellon University (CMU) Pocketsphinx, the state of the art speech recognition engine developed at CMU. Our method is to apply knowledge-based approach for the Arabic sentence structure problem. Certainly, N-Best Arabic syntactic rules are extracted from the tagged Baseline transcription corpus. The extracted rules are then used for rescoring the N-Best hypotheses produced by the ASR decoder.

The paper is organized as follows: In Section 2, we provide a literature review. Sections 3 and 4 introduce data mining approach and the Baseline system, respectively. In section 5, we provide the Arabic phoneme set. Then in Section 6, a description of the Baseline phonetic dictionary is provided. Section 7 describes our methodology followed by Section 8 detailing the testing and evaluation of the proposed method. Then, in section 9, a new design for language models is proposed. Finally, Section 10 presents the conclusion and future work.

## 2 Literature Review

Using linguistic knowledge to improve speech recognition systems was used by many researchers. Salgado-Garza et al. (2004) demonstrated the usefulness of syntactic trigrams in improving the performance of a speech recognizer for Spanish language. Beutler (2007) demonstrated a method to bridge the gap between statistical language models and elaborate linguistic grammars. He introduced precise linguistic knowledge into a medium vocabulary continuous speech recognizer. His results showed a statistically significant improvement of recognition accuracy on a medium vocabulary continuous speech recognition dictation task. Wang et al. (2002) compared the efficacy of a variety of language models (LMs) for rescoring word graphs and N-Best lists generated by a large vocabulary continuous speech recognizer. These LMs differ based on the level of knowledge used (word, lexical features, syntax) and the type of integration of that knowledge. Xiang et al. (2009) presented advanced techniques that improved the performance of IBM Malay-English speech

translation system significantly. They generated linguistics-driven hierarchical rules to enhance the formal syntax-based translation model.

As Arabic Part of speech (PoS) tagging is essential component in our method, we performed the following literature review. The stochastic method dominates PoS tagging models. Diab et al. (2004) presented a Support Vector Machine (SVM) based approach to automatically tag Arabic text. Al-Shamsi and Guessoum (2006) presented a PoS Tagger for Arabic using a Hidden Markov Model (HMM) approach. El-Hadj et al. (2009) presented an Arabic PoS tagger that uses an HMM model to represent the internal linguistic structure of the Arabic sentence. A corpus composed of old texts extracted from books written in the ninth century AD was created. They presented the characteristics of the Arabic language and the set of tags used. Albared et al. (2010) presented an HMM approach to tackle the PoS tagging problem in Arabic. Finally, the Stanford Natural Language Processing Group developed an Arabic tagger (2011) with an accuracy range between 80% and 96%.

According to the literature review, and to the best of our knowledge, we have not found any research work that employs a machine learning algorithm to distill N-Best syntactic rules to be used for rescoring N-Best hypotheses for large vocabulary continuous speech recognition systems.

### 3 Data-Mining Approach (WEKA tool)

WEKA is a collection of machine learning algorithms for data mining tasks which represents a process developed to examine large amounts of data routinely collected. Extracting N-Best syntactic rules using WEKA tool is described in Tobias Scheffer (2005). He presented a fast algorithm that finds the  $n$  best rules which maximize the resulting criterion. The strength of this tool is the ability to find the relationships between tags with no consecutive constraint. For example, if we have a tagged sentence, then it is possible to describe the relations between its tags as follows: if the first word's tag is noun and the sixth word's tag is an adjective, then the ninth word's tag is adverb with certain accuracy. This also could be used for words, i.e. an extracted rule could have  $n$  words with its relationships and accuracy. Data mining is used in most areas where data are collected such as health, marketing,

communications, etc. it worth noting that data mining algorithms require high performance computing machines. For more information about WEKA tool, Please refer to Machine Learning Group at University of Waikato (2011).

### 4 The Baseline System

Our corpus is based on radio and TV news transcription in the MSA. The audio files were recorded from many Arabic TV news channels, a total of 249 business/economics and sports stories (144 by male speakers, 105 by female speakers), with total duration of 7.57 hours of speech. These audio items contain a reasonable set of vocabulary for development and testing the continuous speech recognition system. The recorded speech was divided into 6146 audio files. The length of wave files varies from 0.8 seconds to 15.1 seconds, with an average file length of 4.43 seconds.

The total words in the corpus are 52,714 words, while the vocabulary is 17,236 words. The transcription of the audio files was first prepared using normal non-vocalized text. Then, an automatic vocalization algorithm was used for fast generation of the Arabic diacritics (short vowels). The algorithm for automatic vocalization is described in detail in Elshafei et al. (2006). The Baseline system WER is reported at 16.04%. Alghamdi et al. (2009) has more details of the pronunciation corpus used in this work.

### 5 Arabic Phoneme Set

We used the Arabic phoneme set proposed by Ali et al. (2009) which contains (40 phonemes). This phoneme set is chosen based on the previous experience with Arabic text-to-Speech systems (Elshafei 1991, Alghamdi et al. 2004, Elshafei et al. 2002), and the corresponding phoneme set which is successfully incorporated in the CMU English pronunciation dictionary.

### 6 Arabic Pronunciation Dictionary

Pronunciation dictionaries are essential components of large vocabulary natural language speaker-independent speech recognition systems. For each transcription word, the phonetic dictionary contains its pronunciation in terms of a sequence of phonemes. We used the tool presented

by Ali et al. (2009) to generate a dictionary for the corpus transcription

## 7 The Proposed Method

Rescoring N-Best hypotheses is the basis of our method. The rescoring process is performed for each hypothesis to find the new score. A hypothesis new score is the total number of the hypothesis' rules that are already found in the language syntax rules (extracted from the tagged transcription corpus). The hypothesis with the maximum matched rules will be considered as the best one. Our method can be described using Figure 4.

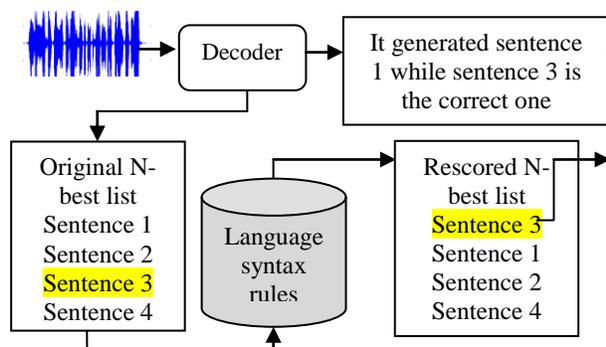


Figure 4. Generation of rescored N-Best list

In Figure 4, suppose that third sentence is the correct sentence that should be returned by the decoder. If the N-Best hypotheses list is rescored using language syntax rules, we expect, hopefully, to get a better result since the final output will be syntactically evaluated. In this case, the hypothesis with maximum number of rules will be chosen since the not-maximum hypothesis is less likely to be the best one. Hence fore, instead of returning the previously top choice (sentence 1) of N-Best list, it will return the top choice of Rescored N-Best list (sentence 3) as shown in Figure 4.

For more clarification, suppose that the two hypotheses of a tested file are as follows:

```
(1) VBD NN NNP DTNNP NN NNP NNP
DTJJ DTNN
(2) VBD NN NNS DTNNP JJ NNP NN DTJJ
DTNNS
```

Each hypothesis will be evaluated by finding the total number of the hypothesis rules that are already found in the language syntax rules.

Suppose that hypothesis number (2) has 4 matching rules while hypothesis number (1) has only 3. In this case, hypothesis number (2) will be chosen as output since it has the maximum matching rules. Since the N-Best hypotheses are sorted according to the acoustic score, if two hypotheses have the same matching rules, the first one will be chosen as it has the highest acoustic score.

Before using WEKA tool, the transcription corpus is tagged using Stanford Arabic tagger which contains 29 tags as shown in Table 1.

#	Tag	Meaning with examples
1	ADJ_NUM	Adjective, Numeric السابع،الرابعة
2	DTJJ	DT + Adjective النفطية،الجديد
3	DTJJR	Adjective, comparative الكبرى،العليا
4	DTNN	DT + Noun, singular or mass المنظمة،العاصمة
5	DTNNP	DT + Proper noun, singular العراق،القاهرة
6	DTNNS	DT + Noun, plural السيارات،الولايات
7	IN	Preposition or subordinating conjunction حرف جر مثل : في حرف مصدرى مثل : أن
...	...	...
29	UNK	Unknown word

Table 1. Stanford tagging set

Finding language syntax rules is performed using a machine learning tool (WEKA-3-6-5). This tool is called to find N-Best syntactic rules. In our method, we choose to find the best 3000 syntactic rules. For more elaboration, Table 2 shows the first best five rules.

1	TAG4=CD TAG6=DTNN 21 ==> TAG5=IN 21 acc:(0.95635)
2	TAG1=VBD TAG3=DTJJ TAG7=DTNN 21 ==> TAG2=DTNN 21 acc:(0.95635)
3	TAG7=CD TAG8=IN 19 ==> TAG9=DTNN 19 acc:(0.95222)
4	TAG7=CD TAG9=DTNN 19 ==> TAG8=IN 19 acc:(0.95222)

Table 2. First 5-Best syntactic rules of the 3000 rules

Our transcription corpus contains sentences that include up to 30 words. So, our rules have the relationships between tags in the range from 1 to 30. The first rule in Table 2 shows that if the fourth word's tag is a number and the sixth word's tag is a noun, then the fifth word's tag will be preposition with rule accuracy of 95.635%. Rule 2 in Table 2 shows the relationships between distant tags (tag1, tag3, tag7, tag2). As example, the following rule provides the relationships between 6 not-consecutive tags.

TAG1=VBD TAG3=DTNN TAG4=DTJJ  
TAG5=NN TAG12=NN ==> TAG2=NN  
acc: (0.92298)

As we mentioned in section 4 that data mining approach to extract association rules in a large data require a high performance computing (HPC) environment. In our experiments, we found that a desktop computer which contains a single processing chip of 3.2GHz and 2.0 GB of RAM can obtain no more than 530 rules. So, extracting high number of rules in a large corpus requires HPC. We used the HPC at KFUPM which described in HPC Center (2011).

## 8 Testing and Evaluation

In order to test our proposed method, we split the audio recordings into two sets: a training set and a testing set. The training set contains around 7 hours of audio while the testing set contains the remaining 0.57 hours. We use the CMU language toolkit to build the Baseline language model from the transcription of the fully diacritized text of 7.57 hours of audio. We used the CMU Pocketsphinx to generate the 50-Best hypotheses and, therefore, to test the proposed method. After intensive investigation of our method, we did not find significant enhancement. However, we found enhancements in some tested files as well as new errors introduced in others. Figure 5 and Figure 6 show enhancement in some tested files.

A waveform of a speech sentence with its text form	 هَذَا وَقَدْ بَلَغَتْ مَبِيعَاتُ شَرِكَةِ فُورْد مُوتُورز فِي الصَّيْنِ خِلَالَ غَامِ أَلْفَيْنِ وَخَمْسَةِ
As recognized by the Baseline	هَذَا وَقَدْ بَلَغَتْ مَبِيعَاتُ شَرِكَةِ فُورْد مُوتُورز التَّسْعِينَ خِلَالَ غَامِ أَلْفَيْنِ وَخَمْسَةِ

system	
Found at →	Hypothesis # 36
As recognized by the enhanced system	هَذَا وَقَدْ بَلَغَتْ مَبِيعَاتُ شَرِكَةِ فُورْد مُوتُورز فِي الصَّيْنِ خِلَالَ غَامِ أَلْفَيْنِ وَخَمْسَةِ

Figure 5. A perfect enhancement in a tested file

A waveform of a speech sentence with its text form	 حَدَرَ الْبَنْكُ الدَّوْلِي دَوْلَ الْخَلِيجِ الْعَرَبِيَّةِ مِنْ صَحْ الْمَزِيدِ مِنْ غَائِدَاتِهَا النَّفْطِيَّةِ فِي مَشْرُوعَاتِ
As recognized by the Baseline system	حَدَرَ الْبَنْكُ الدَّوْلِي دَوْلَ الْخَلِيجِ الْعَرَبِيَّةِ مِنْ صَحْم الْمَزِيدِ مِنْ غَائِدَاتِهَا النَّفْطِيَّةِ فِي مَشْرُوعَاتِ
Found at →	Hypothesis # 50
As recognized by the enhanced system	حَدَرَ الْبَنْكُ الدَّوْلِي دَوْلَ الْخَلِيجِ الْعَرَبِيَّةِ مِنْ صَحْ الْمَزِيدِ مِنْ غَائِدَاتِهَا النَّفْطِيَّةِ فِي مَشْرُوعَاتِ

Figure 6. A perfect enhancement in a tested file

For the tested file in Figure 5 the best hypothesis was found at position #36, while the hypothesis #50 was found to be best one in Figure 6. The previous two examples show a perfect enhancement where a wrong word is switched to a correct one. The following are two other examples to show partial enhancements in the tested files. Figure 7 found the best choice to be the hypothesis #8, while the hypothesis #4 was found to the best one in Figure 8.

A waveform of a speech sentence with its text form	 وَأَكَّدَ التَّقْرِيرُ أَنَّ مُتَوَسِّطَ سِعْرِ السَّلَّةِ فِي شَهْرِ دَيْسَمْبَرِ بَلَغَ ثَمَانِيَةَ وَخَمْسِينَ دُولَارًا وَعَشْرَةَ سِنْتَاتِ
As recognized by the Baseline system	وَأَكَّدَ التَّقْرِيرُ أَنَّ مُتَوَسِّطَ سِعْرِ السَّلَّةِ فِي شَهْرِ السَّنِيوَرَةِ بَلَغَ ثَمَانِيَةَ وَخَمْسِينَ دُولَارًا وَعَشْرَةَ سِنْتَاتِ
Found at →	Hypothesis # 8
As recognized by the enhanced system	وَأَكَّدَ التَّقْرِيرُ أَنَّ مُتَوَسِّطَ سِعْرِ السَّلَّةِ فِي شَهْرِ دَيْسَمْبَرِ بَلَغَ ثَمَانِيَةَ وَخَمْسِينَ دُولَارًا وَعَشْرَةَ سِنْتَاتِ

Figure 7. A partial enhancement in a tested file

A waveform of a speech sentence with its text form	
As recognized by the Baseline system	إِنَّ فِرَقَ الْإِنْتَرِنِتِ
Found at →	Hypothesis # 4
As recognized by the enhanced system	إِنَّ فِرَقَ الْإِنْقَاذِ اللَّهُ

Figure 8. A partial enhancement in a tested file

The previous examples show that our method is a promising method to enhance speech recognition accuracy. However, with enhancements in some tested files, we found new errors (i.e. previously correct recognized words) introduced in some tested files as shown in Figure 9.

A waveform of a speech sentence with its text form	
As recognized by the Baseline system	وَذَلِكَ بِمُشَارَكَةِ عَدَدٍ مِنْ رِجَالِ أَعْمَالٍ وَمُسْتَثْمِرِينَ سُعُودِيِّينَ
Found at →	Hypothesis # 9
As recognized by the enhanced system	وَذَلِكَ بِمُشَارَكَةِ عَدَدٍ لِرِجَالِ أَعْمَالٍ وَمُسْتَثْمِرِينَ سُعُودِيِّينَ

Figure 9. A wrong hypothesis selection example

We also would like to present a case where the N-Best hypotheses already have the correct choice but was not selected after the rescoring process. Figure 10 shows an example.

A waveform of a speech sentence with its text	
---	---

form	السُّعُودِيَّةُ
As recognized by the Baseline system	أَفَادَتِ دِرَاسَةُ حَدِيثُهُ عَنِ التَّمْوِيلِ الْعَقَارِيِّ السُّعُودِيَّةُ
The chosen →	Hypothesis # 4
As recognized by the enhanced system	أَفَادَتِ دِرَاسَةُ حَدِيثُهُ عَنِ التَّمْوِيلِ الْعَقَارِيِّ سُعُودِيَّةُ
The correct →	Hypothesis # 3
Neither Baseline nor enhanced	أَفَادَتِ دِرَاسَةُ حَدِيثُهُ عَنِ التَّمْوِيلِ الْعَقَارِيِّ فِي السُّعُودِيَّةُ

Figure 10. Not-selected correct hypothesis example

In our method, part of speech tagging was crucial to support the correctness of the method used. Even though the Stanford tagger which was used in our method has many correct tagged sentences, however, there are many mistakenly tagged sentences. We provide two examples of a correct tagged sentence and a wrong tagged one as shown in Figure 11.

A correct tagged sentence	
قالت/VBD	أرامكو/NN شركة/NNP
السعودية/DTNNP	دال/NNP وشركة/NN
اليوم/DTNN	الأمريكية/DTNNP كيميكلز/NNP
A wrong tagged sentence	
الجمهورية/DTNN	إن/NN متقى/JJ وقال/NN
أن/NN	على/IN مضممة/DTNN الإسلامية/VN
فعالاً/NN	للفظ/VN مزودا/VBP تكون/NN
وَجَدِيرًا/NN	بِالثقة/JJ

Figure 11. Two examples of tagged sentences

In Figure 11, the highlighted texts were wrongly tagged. So, extracting the language syntax rules using many errors will not be strong enough for rescoring the N-Best hypotheses. This is our justification of our result, enhancement in some tested files and new errors in others.

In addition to the tagger problem, we finalize this section by explaining the effect of diacritics in this research work. Not like English, Arabic sentences are diacritized. Accordingly, the N-Best

hypotheses will be diacritized. Acoustic score also provided for each hypothesis as shows in Figure 12.

9106-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9179-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9320-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9130-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9203-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9344-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9564-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9588-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9609-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9633-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9655-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9679-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9756-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9780-	الَّتِي تَعْتَمِدُ عَلَى الْغَازِ فِي السُّعُودِيَّةِ
9909-	الَّتِي تَعْتَمِدُ عَلَى لِمَصْفَى السُّعُودِيَّةِ

Figure 12. 10-Best list of a tested file.

It is noted that the N-best hypotheses produced by the ASR system are diacritized, which results in many hypotheses that differ only in the diacritics, thus reducing the variety of hypotheses that are included in the N-best list for any value of N. The highlighted hypotheses in Figure 12 are examples. This same-tags case prevents the diversity that should be presented in the N-Best hypotheses. One case, among 300-Best hypothesis, we found 16 different hypotheses, (i.e. at words level). As the acoustic scores are sorted in decreasing order, the problem showed up when, as example, finding the first 50 hypotheses with same words and different diacritics. So, instead of searching among first different hypotheses like English, the search will be away from the high score results, therefore, reducing the accuracy.

## 9 New Designs for Language Models

Even though our method does not increase the Baseline accuracy, it introduces a new design for language models. We propose to relax the constraint of having consecutive few words which usually used to build language models. Cao et al. (2006) demonstrated that many manually identified relationships can be hardly extracted automatically from corpora. This is why they used hand-crafted thesauri (such as WordNet) and co-occurrence relationships for limited relations related to nouns (synonym, hypernym and hyponym). Ruiz-Casado et al. (2007) describes an automatic approach to

identify lexical patterns that represent semantic relationships between concepts in an on-line encyclopedia. They have found general patterns for the hyperonymy, hyponymy, holonymy and meronymy relations. Figure 13 shows our proposed framework. It shows that instead of finding words relations based on specific types, we propose to find words' relations with no restrictions (i.e. in general)

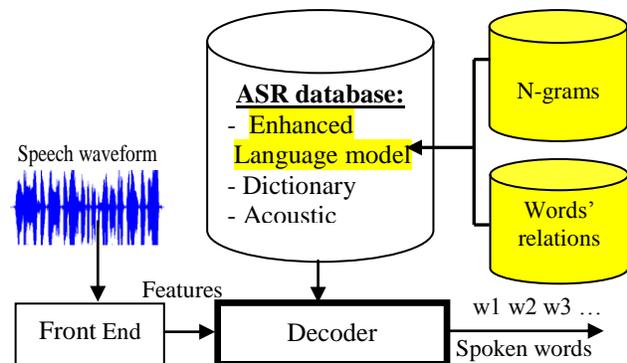


Figure 13. A proposed framework for language models

Figure 13 shows that instead of building the language models based on few consecutive words, the language models could account for longer-distance constrains which we called Enhanced language model. The longer-distance relations have no constraints regard the number of words (such as two or three) or type (such as synonyms). As we mentioned in section 8 (the proposed method) that WEKA tool can extract the relations of many tags. In the same way, we propose to use WEKA to extract the relationships between different words within the same sentence. There are no restrictions of the numbers of words, as the current language models which deal with 3 consecutive words maximum. WEKA tool can generate N-Best rules which can be used as a complement module of the standard language models. In this case, instead on having one module, two modules will be used in computation the words consecutive score. For example, the following cases illustrate how to utilize WEKA tool to extract words' relationships. So, as the rule:

TAG1=VBD TAG3=DTNN TAG4=DTJJ  
TAG5=NN TAG1.2=NN ==> TAG2=NN

We can extract a similar rule but directly with words as follows:

word1=حددت word3=الحج  
word4=السعودية word5=معيار  
word12=المقبل ==> word2=وزارة

In this case, 6 words can contribute to find the best sentence which is better than n-grams which require the words to be executives and usually built using (2-3) words.

## 10 Conclusion and Future Work

In this paper, we conclude that N-Best rescoring for Arabic speech recognition (using Arabic data-driven syntax) does not provide significant enhancement. However, more investigation can be performed with a high accurate part of speech tagging model.

As future work, we recommend to utilize linguistic knowledge at the decoder level, i.e. before releasing the decoder output. We also recommend to do further research on Arabic part of speech tagging, especially for diacritized text.

## Acknowledgments

This work is supported by Saudi Arabia Government research grant NSTP # (08-INF100-4). The authors would like also to thank King Fahd University of Petroleum and Minerals for its support of this research work.

## References

- Bing Xiang, Bowen Zhou and Martin Cmejrek. 2009. Advances in syntax-based Malay-English speech translation. Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE Computer Society: 4801-4804.
- Dan Jurafsky and Martin J. 2009. Speech and Language Processing, second edition, Pearson.
- Fatma Al-Shamsi and Ahmed Guessoum. 2006. A Hidden Markov Model –Based PS Tagger for Arabic, CiteSeerX.
- Guihong Cao, Jian-Yun Nie, and Jing Bai. 2005. Integrating word relationships into language models. Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. Salvador, Brazil, ACM: 298-305.
- High Performance Computing (HPC) Center, 2011. <http://hpc.kfupm.edu.sa/Home.htm>
- Luis R. Salgado-Garza, Richard M. Stern and Juan A. Nolasco F. 2004. N-Best List Rescoring Using Syntactic Trigrams, MICAI 2004: Advances in Artificial Intelligence.
- Machine Learning Group at University of Waikato, 2011. <http://www.cs.waikato.ac.nz/ml/WEKA/>
- Mansour Alghamdi , Moustafa Elshafei , and Husni Almuhtasib . 2009. Arabic broadcast news transcription system, International journal of speech and technology, 10: 183–195
- Mansour Alghamdi , Husni Almuhtasib, and Moustafa Elshafei . 2004. Arabic Phonological Rules, King Saud University Journal: Computer Sciences and Information. Vol. 16, pp. 1-25
- Maria Ruiz-casado , Enrique Alfonseca , and Pablo Castells. 2007. Automatising the learning of lexical patterns: an application to the enrichment of WordNet by extracting semantic relationships from Wikipedia. Data Knowledge and Engineering, 61 3 , pp. 484–499.
- Mohammed Albared, Nazlia Omar, Mohd.Aziz, and Mohd Ahmad Nazri. 2010. Automatic part of speec tagging for Arabic: an experiment using Bigram hidden Markov model, RSKT'10 Proceedings of the 5th international conference on Rough set and knowledge technology
- Mohamed Ali, Moustafa Elshafei, Mansour Alghamdi, Husni Almuhtaseb and Atef Alnajjar, 2009. Arabic Phonetic Dictionaries for Speech Recognition. Journal of Information Technology Research, Volume 2, Issue 4, pp. 67-80.
- Mona Diab, Kadri Hacıoglu, and Daniel Jurafsky. 2004. Automatic tagging of Arabic text: from raw text to base phrase chunks, 5th Meeting of the North American Chapter of the Association for Computational Linguistics/Human Language Technologies Conference.
- Moustafa Elshafei. 1991. Toward an Arabic Text-to-Speech System, The Arabian Journal of Science and Engineering, Vol. 16, No. 4B, pp.565-583.
- Moustafa Elshafei , Husni Almuhtasib and Mansour Alghamdi. 2002. Techniques for High Quality Text-to-speech, Information Science, 140 (3-4) 255-267.
- Moustafa Elshafei., Husni Al-Muhtaseb, and Mansour Alghamdi. 2006 . Machine generation of Arabic diacritical marks. In Proceedings of the 2006 international conference on machine learning; models, technologies, and applications (MLMTA'06), USA.
- René Beutler. 2007. Improving Speech Recognition through Linguistic Knowledge, Doctoral Dissertation ,ETH Zurich.
- Stanford Log-linear Part-Of-Speech Tagger, 2011. <http://nlp.stanford.edu/software/tagger.shtml>
- Tobias Scheffer . 2005. Finding association rules that trade support optimally against confidence. Intell. Data Anal. 9(4): 381-3
- Wen Wang, Yang Liu, and Mary P. Harper. 2002. Rescoring effectiveness of language models using different levels of knowledge and their integration. Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on.
- Yahya El Hadj, Imad Abdulrahman Al-Sughayeir and Abdullah Mahdi Al-Ansari. 2009. Arabic Part-Of-Speech Tagging using the Sentence Structure, Proceedings of the Second International Conference on Arabic Language Resources and Tools, The MEDAR Consortium.