10 June 2019

Name: _Ghribi Saif Eddine_**(not in pair)**

Symbolic Natural Language Processing (SD213)

Project Topic: Create a grammar for the Tunisian Arabic

1. Introduction

The Arabic language is the native language of over 200 million speakers and it is totally spoken by over 450 million, which makes it the fifth most spoken language in the world. In fact, the Arabic, like Hebrew, is a Semitic language and it is always written from the right to the left. In almost every Arabic speaking country we find two kinds of Arabic, the standard which is usually the official language. It is also the one that everyone uses to write documents and it is common for all these countries . Unlike the spoken Arabic that is specific to every country and the variations are due to especially historical events like the language of a colonizer or the language of countries traders used to visit frequently. The Tunisian Arabic can be defined as a dialect of Maghrebi Arabic spoken by the Tunisians who are over 11 million speakers. Being no exception, this dialect's vocabulary contains mostly Arabic but we can find words taken from French,Turkish,Italian and Spanish. This has a raised a need within young people especially with the rise of the texting era in the last two decades. So Tunisian youth who were used to write in French, found writing with Arabic keyboard hard and it is very weird to write French words with Arabic. Thus, they still write Arabic with the Latin alphabet but with adding numbers to cover all the phonetics of the Arabic. These numbers are :5 -> Khā,7 -> Ḥā,9 -> Qāf ,8 -> Ġayn ,3 -> ʿAyn.

This syntax is the one we will use throughout this project since it is the most used by Tunisians.

So this is an example of the difference between the Arabic and the Tunisian one:

<p style="text-align:center;color:orange">N7eb(I love) esfar(traveling)</p>

<p style="text-align:center">أحب السفر</p>

<p style="text-align:center">I love traveling.</p>

Here, we can notice there are many problems that our grammar needs to be aware of and must understand.

2. Problem

The problems found in the last example are mostly:

The personal pronoun is absent in both Tunisian and standard Arabic, this is very similar to Spanish so we notice that sentences can start with a verb.

The grammar should also understands that some verbs in Arabic in general require a specific coherence like for example one cannot eat a house.

3. Claim

Thus ,In this project, we will try to emphasize and explore the difference between the classical Arabic and the Tunisian Arabic. Thanks to techniques learnt during the course of Symbolic natural language processing we will create a small grammar with prolog to understand more how we can speak the Tunisian Arabic.

4. **Relevant Studies**

The Wikipedia article on the Tunisian Arabic is considered as a research that can be found in the famous research gate platform and it was written by Tunisian researchers.

In addition, a valuable study about Arabic entitled "Formal Description of Arabic Syntactic Structure in the Framework of the Government and Binding Theory" is to be explored also since DCG rules have been developed for Arabic in this study.

5. **Discussion:**

Like Arabic ,Tunisian version has two types of sentences ,verbal and nominal ,verbal are the most common but nominal are not to be neglected .

For the verbal phrases,we will be limited to 3 rules in our grammer.

```
vp --> v(_,_).
vp --> v(Sem1,Sem2),np(Sem1,Sem2). % Sem1 (Semantic feature 1) , Sem2 (Semantic feature 2)
vp --> v(Sem1,Sem2),pp(Sem1,Sem2).
```

- In tunisian Arabic we can have a sentence that is formed by a verb only like an answer to the question: Are you in work ?

Answer: rawa7t.

I returned (home).

- There is also the famous verbal phrase structure which is a verb and a nominal phrase such as:

Chrit(I bought) baytha(egg) sghira(small).

I bought a small egg.

- And the final form is to have a verb plus a prepositional phrase such as :

Mchit(I went) lel(to the) dar(house)

I went to the house

For the nominal phrases we have 4 types:

```
np --> nn(Numb,_,_),adj(Numb).
% type 1
np(Sem1,Sem2) --> nn(_,Sem1,Sem2).
% type 2
np(Sem1,Sem2) --> nn(Numb,Sem1,Sem2),adj(Numb).
% type 3
np(Sem1,Sem2) --> cp(Numb),nn(Numb,Sem1,Sem2). % Numb for plural or singular
% type 4
np(Sem1,Sem2) -->nn(_,Sem1,Sem2), conj, nn(_,Sem1,Sem2).
```

The first line shows that we can say that a noun plus an adjective is considered as a nominal phrase in Arabic and the other ones are to be used within the verbal phrases.

Examples for the 4 types:

Type	Example
1	Sou9 (the souk,the market)
2	Khouya sghir(my little brother)
3	Barcha jus(a lot of juice)
4	Khouya w bouya (my brother and my father)

To create this grammar, semantic features are necessary. Thus, the table below summarizes the semantic features illustrated with examples that are so frequent in the everyday spoken Tunisian Arabic:

Nature	Semantic Feature1	Semantic Feature 2	Semantic Feature 3	Example in Tunisian Arabic	Translation
Verb	Edible	Non drinkable		Klit	I ate
	Edible	Drinkable		Chrabet	I drank
	Location	_	_	Mchit	I went
	Sentient being	_		Tlabt	I called
	Edible	_		Chrit	I bought
Noun	Singular	location	_	Dar	house
	Plural	Location	_	Diar	houses
	Singular	Sentient being	_	Sa7bi	My friend
	Plural	Sentient being	_	S7abi	My friends
	Singular	Edible	Non drinkable	Baytha	An egg
	Plural	Edible	Non drinkable	Bayth	eggs
	Plural	Edible	Drinkable	Jus	juice
Conjunction	_			W	and
Preposition	location	_		Lel	To(toward)
Adjective	Singular	_		Sghir	Small
	Plural	_		Sghar	Small

These are some sentences generated by our grammar:

```
6 ?- vp(['a3tini',A,B],[]).
A = baytha,
B = sghira ;
```

A3tini baytha sghira

Give me a small egg

```
12 ?- vp(['chrabet',A],[]).
A = jus ;
A = ma2 ;
```

<div align="center">

Chrabt ma2

I drank water

```
13 ?- vp(['tlabt','bouya','w','khouya'],[]).
true .
```

Tlabt bouya w khouya

I called my father and my brother

</div>

6. Conclusion

Thanks to this grammar, (source code is joint as grammar.pl) we managed to generate 487 correct Tunisian Arabic sentences that are frequently used. After accomplishing this project, we can conclude that one can develop a solid SNLP application for a dialect. Nonetheless, even if it is a dialect, the rules of the mother language can be still spotted which means we need to add all the rules of the Arabic language, which is an interesting project.

7. Bibliography

https://www.vistawide.com/arabic/arabic.htm
https://en.wikipedia.org/wiki/Tunisian_Arabic
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-55462014000300015