# CHAPTER 1

# INDTRODUCTION

This chapter includes the following subtopics, namely: (1) Rationale; (2) Theoretical Framework; (3) Conceptual Framework/Paradigm; (4) Statement of the problem; (5) Hypothesis (Optional); (6) Assumption (Optional); (7) Scope and Delimitation; and (8) Importance of the study.

## 1.1 Rationale

Indonesia is a country where most followers of the religion are Islam. About 232 million people are Muslims. Muslims have a holy book, the Qur'an. A written document or text that is more or less 15 centuries ago in which Muslims are obliged to do what is ordered and stay away from what is prohibited written in the Qur'an. Al-Qur'an is written in Arabic so that many Indonesian people who learn Arabic to be able to understand the contents of the Qur'an. Processing Arabic becomes necessary to analyze the Qur'an further or make tools to facilitate learning Arabic. Before doing so back to basics of Linguistics is a morphological analysis..

In Arabic, the science of morphology is called *Sharaf*. *Sharaf* is the basis of Arabic and is also called the science of tools to understand sentences in Arabic and the key to opening a repository of Islamic knowledge. An important part of morphological is the formation of sentences or classes of words. Different from other languages, for example, in Indonesian, there are 13 classes of words. In Arabic, there are only 3, i.e., *isim* (noun), *fi'il* (verb) and *huruf* (particle). Not that other words are omitted, it's just examples such as adjectives, adverbs, pronouns that belong to the *isim* group. While prepositions, conjunctions, question words enter into *huruf*. So the three elements above are the root or core of all the word classes that exist[21]. Whereas all verbs include *fi'il* but not all *fi'il* are verbs. Several adjectives fall into the *fi'il* category [3].

In Sharaf (morphology), the most important thing is to regulate or focus on changing the form of words to other words or commonly referred to as *tashrif* [3]. The pattern of word formation or commonly called *wazan* is created. *Tashrif* divided into two, namely *tashrif lughawi* (inflection) and *tashrif istilahy* (derivation) [33]. In Arabic, the changes in word types include types (*fi'il* / *isim*), *dhomir* / pronoun and *wazan* or patterns of the formation of these words. The changes and patterns of each word to another influence the semantics or meaning of the word. Therefore it would be better if there is a system or program that can identify the word both from inflection and its derivation for morphological

analysis of the Qur'an.

The Gonzales paper [27], a reference paper from this study, has conducted a morphological analysis of Arabic. Jabalin application that can be accessed on the site `http://elvira.lllf.uam.es/jabalin/analizarForma.php` has been able to identify and generate the infected word and its derivation successfully. The paper only focuses on one element of the word that is *fi'il* or verb. That study explicitly said that the future work is to work on morphological analysis of *isim* or nouns. Therefore, this research focus on the identification model. In addition to only being able to be identified, it can also identify *isim*. But here *isim* is limited to the only isim whose root is derived from the verb.

In the application of the annotation of the Qur'an that is currently an identification of the types of words and *dhomir* (pronoun), but there is no pattern / *wazan*, which is where this pattern is essential to know the meaning and or changes in other words of the word. The Gonzales paper [27] focuses more on *fi'il* or verbs, while in this study the *isim* or noun is added which is the future work in the paper. From these two things, this research focuses on identifying patterns and also identifying *isim*. It's just that *isim* in this study focus on the *isim* whose roots come from *fi'il*.

The current state of the art for Arabic morphological analysis is rule-based. However, this study will be using a neural-based approach or using the deep learning method. The method, in this case, is a recurrent neural network because in the current morphological task paper, which is SIGMORPHON 2018 [14] mainly using neural-based, and the result is good. Try to implement that in this study to help to make a morphological analyzer model.

## 1.2 Theoretical Framework

This system input will be the Arabic word, and the output is its MSD, or morphosyntactic description [44], or morphological features [24]. On the Tabel 1.1 is the example of

Table 1.1: Input Example from 3 different languages

| No | Language | Input (Word) | Output (MSD) |
|---|---|---|---|
| 1 | English | Run | pos=V,mood=IND,tense=PST,per=3,num=SG |
| 2 | Indonesia | makan | pos=V,voice=ACT |
| 3 | Arabic | يَعْلَمُوْنَ | pos=V,mood=IND,aspect=IPFV,voice=ACT, GEN=M,PER=3,NUM=PL,verb-form=Iia |

input and msd from 3 different common languages, and the Arabic got more MSDs, so that makes the Arabic word is interesting to study. Although on the Arabic side, the MSD can

be classified into three namely types of word, pronoun(*dhamir*), and verb form (*wazan*), and will be discussed more later on (See on Table 1.2).

Table 1.2: Arabic side of MSDs

| No | Arabic | MSDs |
|---|---|---|
| 1 | Type of words (madhi,mudhari,amr,nahyi) | pos={V/N},mood={IND/IMP}, aspect={IPFV/PFV},voice={ACT/PASS} |
| 2 | Dhamir (pronouns) | GEN={M/F},PER={1/2/3}, NUM={SG/DU/PL} |
| 3 | Wazan | verb-form={1-22} |

## 1.3 Conceptual Framework/Paradigm

The previous recent study is Gonzales paper [27] there is no Noun tag identification. On the recent morphological world contest, Sigmorphon 2018 [14] there is no verb form or *wazan* identification which *wazan* is important on the Arabic word itself according to book [3]. So in this study is for focussing on fulfilling weaknesses from these two related studies, which are adding noun tag and add verb form identification and wrap it to msd form.

Table 1.3: Dictionary dependency problem example that the two type of word must be known before both of it

| Fiil Madhi | Fiil Mudhari | Wazan |
|---|---|---|
| ضَرَبَ | يَضْرِبُ | Iai |
| ضَرَبَ | x | ? |
| x | يَضْرِبُ | ? |

To identify the verb form, especially on form Iau till Iii (form I), the two types of word *fiil madhi* and *fiil mudhari* must be known before for both of it to know its verb form 1.3. To know the type of word of the word, you must see the Arabic dictionary, but the resource or the API is limited, so it must identify its verb form without using the dictionary. It is important because form Iau and Iii is the most common verb form on Arabic according to the Holy Qur'an.

## 1.4 Statement of the Problem

According to the problem explanation and the weaknesses from the previous study, the problem is how to identify the verb form without a dictionary and how to add a noun to the data and wrap it into the msd form.

## 1.5 Objective and Hypotheses

The purpose of this study is the same with aim morphological analysis of the Arabic language. The objective is to establish the formalization of the words in Arabic itself. Try to add some other data type of word, which is a noun, to the dataset. Using a deep learning approach ( recurrent neural network ) hope can help identify the MSD of the Arabic word better. The recurrent neural network can capture the information of the sub of the word. Neural-based also can eliminate the dictionary dependency problem. It is just read from all the data and see the sub-word sequence connection and mapping to the verb form to see the result.

## 1.6 Assumption

Using RNN (Neural method approach) that can capture sequence information prefix, infix, and suffix of a word can identify MSD with noun type of word and improve Jabalin performance.

## 1.7 Scope and Delimitation

The Arabic word in this study follow the rules, namely :

1. One Arabic with full of diacritics (*harakat*)
2. The Arabic word must be in the normal form not affected by the sentences rule (*nahwu*)
3. Active word only
4. The Arabic word roots do not contain weak letter which are ا, و,ي
5. The noun word is noun which derived from a verb

## 1.8 Significance of the Study

This study can help provide some insight into some of the linguistic elements at work in grammar, especially in the Arabic language. This study also helps to improve the last gap that the previous method can not execute from the previous method ( morphological analyzer on POS noun ). Finally, this study also enhances what can be identified from the previous approach to better perform with this study method.