



**ON CREATING A MACHINE READABLE LINGUISTIC RESOURCE
PROTOTYPE FOR AUTOMATIC RETRIEVAL OF INDONESIAN
VERBS**

A THESIS
In Partial Fulfillment of the Requirements
For the Sarjana Degree Majoring Linguistics in English Department
Faculty of Humanities Diponegoro University

Submitted by:
StevaniApriliaRuswati
13020112120008

**FACULTY OF HUMANITIES
DIPONEGORO UNIVERSITY
SEMARANG 2017**

PRONOUNCEMENT

The writer honestly confirms that she complies this thesis by herself and without taking any result from other researcher in S-1, S-2, S-3 and in diploma degree of any university. The writer ascertains also that she does not quote any material from publications or someone's paper other than from the references mentioned.

Semarang, August 2017

StevaniApriliaRuswati

MOTTO AND DEDICATION

Desire without knowledge is not good—how much more will hasty feet miss the way!

-PROVERBS 19:2

*No matter what storm you face, you need to know that God loves you.
He has not abandoned you.*

-FRANKLIN GRAHAM

*This thesis is dedicated to
My beloved Mom, Dad, my little family, my sister, and
to everyone who had contributed in completing this thesis.*

**ON CREATING A MACHINE READABLE LINGUISTIC RESOURCE
PROTOTYPE FOR AUTOMATIC RETRIEVAL OF INDONESIAN
VERBS**

Written by

StevaniApriliaRuswati

NIM: 13020112120008

is approved by the thesis advisor

On August 24th, 2017

Thesis Advisor

Prihantoro, S.S., M.A

NIP. 198306292006041002

The Head of the English Department

Dr.AgusSubiyanto, M.A.

NIP. 196408141990011001

VALIDATION

Approved by

Strata 1 Thesis Examination Committee

Faculty of Humanities Diponegoro University

On October 12th, 2017

Chair Person

First Member

Dr. Nurhayati, M.Hum
NIP. 196610041990012001

Dr. Y. Herudjati Purwoko, M.Sc
NIP. 195303271981031006

Second Member

Third Member

Ayu Ida Savitri, SS, M.Hum
NIP. 19790822200812013

Dra. R. Aj. Atrinawati, M.Hum
NIP. 196101011990012001

ACKNOWLEDGEMENT

Praise to Almighty God, who always blesses and gives me strength to complete this thesis entitled “On Creating a Machine Readable Linguistic Resource Prototype for Automatic Retrieval of Indonesian Verbs”. Foremost, I would like to express my sincere gratitude to my advisor, Prihantoro, S.S, M.A. for the continuous support of my research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me all the time of research and writing of this thesis.

My deepest thank also goes to some people in the following:

1. Dr. Redyanto M. Noor, M.Hum., as the Dean of Faculty of Humanities, Diponegoro University.
2. Dr. Agus Subiyanto, M.A., as the Head of English Department, Faculty of Humanities, Diponegoro University.
3. All of the lectures in English Department, particularly in Linguistics section Faculty of Humanities, Diponegoro University.
4. My beloved parents, Rustomo, S.PAK and Sriwati who always support me in accomplishing this thesis in their prayers.
5. My beloved family, Jefri Mokundu, my husband and Christabel Andira Mokundu, my little daughter who always cheer me up.
6. My sister, Gusvinanda Ellya Ruswati who never stopped asking me about my thesis progress.

7. My best college friends Widi and Radiani. They helped and supported me during college activities.
8. Linguistics classmate, especially Dianika, Diana and Naufal who always shared knowledge with me, helped me when I have not understood the lesson yet and all of my friends in English Department 2012.
9. All people who have contributed finishing this thesis who I cannot mention one by one.

I realize that this thesis is still far from being perfect. Therefore, I will be glad to receive any constructive criticism and recommendation to make this thesis better.

Finally, I expect that this thesis will be useful to the reader who wants to learn how to create automatic retrieval in certain linguistic application.

Semarang, September 2017

Stevani Aprilia Ruswati

TABLE OF CONTENT

TITLE.....	i
PRONOUNCEMENT	ii
MOTTO AND DEDICATION	iii
APPROVAL.....	iv
VALIDATION	v
ACKNOWLEDGEMENT	vi
TABLE OF CONTENT	viii
ABSTRAK	xi
CHAPTER I INTRODUCTION	1
1.1 Background of the Study	1
1.2 Research Problems	2
1.3 Purposes of the Study	3
1.4 Scope of the Study.....	3
1.5 Previous Study.....	4
1.6 Organization of the Writing	5
CHAPTER IILITERATURE REVIEW.....	6
2.1 Complexities of Prefixes	6
2.2 Ambiguities	12

2.3.Morphophonemic	13
2.3.2 Morphophonemic Process	14
2.4 Machine Readable Linguistic Resource	15
CHAPTER III RESEARCH METHODOLOGY	17
3.1 Data	17
3.2 Procedures	18
3.2.1 Collecting Wordlist.....	18
3.2.2 Reviewing Literature	19
3.2.3 Analysing Data	19
3.3 Tags.....	19
CHAPTER IV ANALYSIS	Error! Bookmark not defined.
4.1 Dictionary.....	Error! Bookmark not defined.
4.1.1 Raw Dictionary	Error! Bookmark not defined.
4.1.2 Enhanced Dictionary.....	Error! Bookmark not defined.
4.2 Application on Corpus Data	Error! Bookmark not defined.
4.2.1 Text Annotation Structure (TAS)	Error! Bookmark not defined.
4.2.2 Retrieval.....	Error! Bookmark not defined.
CHAPTER V CONCLUSION.....	Error! Bookmark not defined.
References.....	Error! Bookmark not defined.

ABSTRAK

Dalam melakukan penelitian, para linguist dapat memanfaatkan teknologi komputer untuk menunjang penelitiannya. Dalam skripsi ini, penulis akan menunjukkan peran teknologi yang berbentuk aplikasi *machine readable linguistic resources*, yaitu NooJ Application, dalam penelitian, terutama saat melakukan analisis data. Tujuan penulis skripsi ini adalah untuk menjelaskan bagaimana membuat sumber, menjelaskan bagaimana menerapkan sumber, serta mendeskripsikan bagaimana mengambil informasi dari korpus. Data yang digunakan adalah nomina yang sering digunakan yang diambil dari Kamus Besar Bahasa Indonesia dengan menggunakan teknik stropektif yang kemudian akan diderivasikan ke verba dengan menggunakan prefik pembentuk verba *m eN-*, *beR-*, *teR-*, dan *dandi-*, sedangkan sumber datanya adalah 25 karya ilmiah linguistik yang telah diakreditasi oleh Dikti. Hasil analisis menunjukkan bahwa dalam menggunakan aplikasi data korpus, dibutuhkan struktur anotasi teks. Struktur anotasi teks ini akan sangat berperan dalam melakukan penelusuran atau penggalan informasi. Ada tiga metode dalam penelusuran, yaitu berdasarkan ortografi, berdasarkan tata bahasa dan berdasarkan semantik.

Keywords: resources, corpus data, machine readable linguistic resources, noun, verb, derivation

CHAPTER I

INTRODUCTION

1.1 Background of the Study

In this globalisation era, technology takes important roles in our life. The benefit of using technology as the main instrument in our life is to make our works more effective and efficient.

One of technology products is computer. It keeps changing and developing year by year. In the beginning, computers can be only used for computing function. However, today, we can type by using computer, as well. People used to type using machine in writing reports, papers, and others. Typing becomes faster by computer.

The computer can help us doing experiments and also researches. It depends on the users' needs on computer. In linguistics case, the computer can help the researchers in doing data analysis.

One of the linguistic products is corpus. Corpus contains resources that linguist need. There are some corpus resources, such as BNC (Aston, 1998), COCA (Davies, 2009), ICNALE (Ishikawa, 2013) which are available online. They have provided their own interfaces, their own corpus and the tools to process the corpus. In applying these, we must access the websites and the corpus is ready to be analysed. We do not need to download any tool or any corpus processing software. However, we cannot modify the corpus data and

we cannot analyse our own corpus. We can only analyse the corpus that they have.

If we want to analyse our own corpus, we can apply corpus processing software, such as NooJ (Siberztein, 2003), AntConc (Anthony, 2004), Unitex (Paumier, 2013) etc. We should download the software and provide the corpus to do an analysis where the benefit is that we can analyse data that we need. In this research, I used NooJ Application as the tool for analysing corpus. NooJ provides devices to explain inflectional and derivational morphology, dictionary, local grammars, syntax and semantic (Siberztein, 2003).

According to the phenomenon mentioned above, in this paper, I will discuss how to use technology for research, especially research on corpus. I will explain how to use corpus processing software to retrieve words in the corpus data, not only orthographically but also grammatically, such as by its affixes.

1.2 Research Problems

In performing this research, there are four research problems that will be discussed and answered in this paper:

- (1) How to make those resources?
- (2) How to apply the resources?
- (3) How to retrieve corpus data by using the corpus processing software?

1.3 Purposes of the Study

Based on the research problems, the purposes of this study are

- (1) To demonstrate how to make the resources.
- (2) To explain the application of the resources.
- (3) To inform how to retrieve information from the corpus.

1.4 Scope of the Study

To make this research specific, I have made some limitations for the study.

I will focus on the specific data, word list and grammar. For the corpus data, I decided to use Indonesian linguistic journals. There are two kinds of journals based on the accreditation: accredited journal and unaccredited journal. Here, I chose to collect data from Dikti accredited journal. I chose them because journal that are accredited by Dikti are more qualified and scientific.

The word list which would be used to make the resources contains frequently used nouns. While, for the grammar, I decided to focus on prefixes which their derivation results are only verbs. The verbalizer prefixes I study here are *meN-*, *beR-*, *teR-* and *di-*. I decided to use those verbalizer prefixes because they are the foundation or the basic of this research.

1.5 Previous Study

This kind of research has been done by other researchers. However, complete previous studies will be discussed further in the literature review section. Several relevant results are discussed briefly in this part.

Prihantoro (2014), in his paper, discusses retrieval method in using machine readable grammar. In his research, he compared the performance of automatic retrieval in a text corpus by regular expression and Local Grammar Graphs (LGG). The result of his research is the preference of retrieval methods depends on the goal of retrieval. In my research is that I focus on using regular expression in retrieving words.

Another research was done by Sarosa (2005). He wrote a morphophonemic in Indonesian affixation. The goal of his research is describing the morphophonemic process on affixation process of Indonesian. In his analysis, he used basic changes of morphophonemic: phoneme change, phoneme loss, phoneme addition and phoneme shift. The result of his study is that there are fourteen variations of morphophonemic process. The difference with my study is that I focus on prefix only: meN-, beR-, teR-, and di-, while he analysed affixes: prefixes, infixes and suffixes.

1.6 Organization of the Writing

To make this paper organized, this research is ordered as follow:

CHAPTER I INTRODUCTION

This chapter includes background of the study, research problems, purpose of the study, scope of the study, previous study, and writing organization.

CHAPTER II LITERATURE REVIEW

This includes the supporting theories of this research and is very helpful in conducting data analysing.

CHAPTER III RESEARCH METHODOLOGY

In this chapter, I will present the corpus data, procedure, and tagging methodology.

CHAPTER IV DATA ANALYSIS

It is the central section of this thesis. It presents the process of making dictionary which includes the raw dictionary and enhanced dictionary and application on corpus data which includes the text annotation structure and the retrieval.

CHAPTER V CONCLUSION

It presents the conclusions of the research.

CHAPTER II

LITERATURE REVIEW

In this chapter, I will discuss some relevant studies that are helpful in conducting this research. The related studies are about the complexities of prefixes in Indonesian, ambiguities, and machine readable linguistic resource.

2.1 Complexities of Prefixes

In constructing words, there are several strategies: affixes, reduplication, combination, acronym, and conversion. In this research, I focus on constructing words by affixes. There are several kinds of affixes. They are prefix, confix, suffix and infix. In this research, I will focus on prefixes as the initial part or research and the other types of affixes will be analysed on future research.

In the process of building words, affixes can be performed as inflection or derivation. Inflectional affix is an affix that functions as grammatical category statement. Meanwhile, derivational affix is affix that diverts the base forms to the different classes (Parera 2010:24). Derivtion affix is also affix that can change the arguments.

There are several prefixes in Indonesian, they are *ber-*, *me-*, *di-*, *ter-*, *ke-*, and *se-* (Chaer 2008:27). Meanwhile, the prefixes that could form verb are *ber-*, *me-*, *di-*, *ter-*, and *ke-* (Chaer 2008:106).

Similar to Parera (2010), Chaer explains that prefixes could act as inflection or derivation. However, he believes that Prefix *beR-* could only act as derivation. It derives nouns to verbs. The grammatical meanings are in table 2.1.

No	Grammatical Meanings	Examples
1	to have (base) or there is/are (base)	<i>berayah</i> 'to have father'
2	to wear or use (base)	<i>berpita</i> 'to wear ribbon'
3	to ride (base)	<i>bersepeda</i> 'to ride bicycle'
4	to contain (base)	<i>beracun</i> 'to contain poison'
5	to expel or produce (base)	<i>berdarah</i> 'to expel blood'
6	to work on (base)	<i>berladang</i> 'to work on field'
7	to do (base)	<i>berdiskusi</i> 'to do discussion'
8	to call (base)	<i>berkagak</i> 'to call brother/sister'
9	to give (base)	<i>berkhotbah</i> 'to give sermon'

Table 2.1 The Grammatical Meanings of Derivational Prefix *beR-*

In other hand, Chaer (2008) explains that prefix *me-* could act as inflective or derivative prefixes. The different is that inflection *me-* could be replaced by either inflection *di-* or inflection *ter-*. Meanwhile, derivation *me-* could not be replaced by neither *di-* nor *ter-*.

NO	meN-	di- and teR-
1	<i>membaca,</i>	<i>dibaca, terbaca</i>
2	<i>Mengaduk</i>	<i>diaduk, teraduk</i>
3	<i>membatu,</i>	<i>*dibatu, *terbatu</i>
4	<i>Menepi</i>	<i>*ditepi, *tertepi</i>

Table 2.2 The Transformation of Prefix *meN-*

The grammatical meanings of *meN-* are as follow:

No	Grammatical meanings	Examples
1	to eat, to drink, to suck	<i>menyoto</i> 'to eat soto'
2	to expel	<i>mengeong</i> 'to expel sound <i>ngeong</i> '
3	to become	<i>membatu</i> 'to become rock'
4	to go toward	<i>menepi</i> 'to go toward sidewalk'

Table 2.3 The Grammatical Meanings of Derivational Prefix *meN-*

Similar to prefix *meN-*, prefix *teR-* could act either inflectional or derivational prefix. The grammatical meanings of inflectional *teR-* are:

No	Grammatical Meaning	Examples
1	to be able to (base)	<i>Terbaca</i> 'to be able to read'
2	accidentally	<i>Terangkut</i> 'accidentally loaded'
3	Have been happened	<i>Terbakar</i> 'have been burned'

Table 2.4 The Grammatical Meanings of Inflectional Prefix *teR-*

The grammatical meanings of derivational prefix *teR* are:-

No	Grammatical Meanings	Examples
1	the one who is (base)	<i>terdakwa</i> 'the one is charged'
2	the most	<i>terpanjang</i> 'the longest'
3	in condition of	<i>terbalik</i> 'in condition of inverted'
4	suddenly happen	<i>terjerembab</i> 'suddenly fall'

Table 2.5 The Grammatical Meaning of Derivational Prefix *teR-*

Prefix *di-* could act as inflectional and derivational prefixes, as well. Inflectional *di-* prefix is for passive verbs. Their grammatical meanings are the inversion from the active verbs of prefix *me-*. Meanwhile, derivative *di-* only

applies on one stem (*maksud*) which forms *dimaksud*. The writer of the book does not explain why there is only one data.

In contrast to Chaer, Moeliono (1993) does not divide the prefixes and the meaning based on the inflection and derivation. He explains that verbs may be derived by affixes which are prefixes, suffixes, confixes, and infixes the most unproductive (Moeliono 1993:81). He also mentioned that Indonesian verbalizer prefixes *meng-*, *per-*, and *ber-*, and also *di-* and *ter-* which can replace *meng-* on certain clause or sentence.

As it has been commented previously, there are obligatory prefixes that derive verbs from certain verb entries. Free morpheme *darat* ‘land’ and *layar* ‘sail’ must take prefixes *meng-* and *ber-* to derive their status from noun to verb.

Nouns	Verbs	Examples
<i>darat</i> (N)	<i>mendarat</i> (V)	* <i>Pesawatitutelahdarat.</i> <i>Pesawatitutelahmendarat.</i>
<i>layar</i> (N)	<i>berlayar</i> (V)	* <i>Perahusedanglayar.</i> <i>Perahusedangberlayar.</i>

Table 2.6 Verbalizer Prefixes

Furthermore, Moeliono shows that prefixes could act as intransitive verb. The first is intransitive verb with prefix *meng-*. He says that most of intransitive

derivational verbs are derived from nouns and adjectives (Moeliono 1993: 9). For example,

To become <base>

batu 'stone' → *membatu* 'become stone'

To function as/like <base>

supir 'driver' → *menyupir* 'function as driver'

To eat/drink <base>

kopi 'coffee' → *mengopi* 'to drink coffee'

To go to <base>

tepi 'side' → *menepi* 'to go to the side'

To look for <base>

rumput 'grass' → *merumput* 'to look for grass'

To expel sound of <base>

raung 'roar' → *meraung* 'to roar'

The second one is intransitive verbs with prefix *ber-*. They are

To have <base>

atap 'roof' → *beratap* 'to have roof'

To use <base>

bedak 'powder' → *berbedak* 'to use powder'

To expel <base>

telur 'egg' → *bertelur* 'to expell egg'

The next is intransitive derivation verbs with prefix *ter-*. Alwi said that its process is less productive than from the verb itself. However, the general meaning from the derivational verbs with prefix *ter-* is in condition of <base>.

2.2 Ambiguities

Generally, ambiguous means there are more than one analysis. It can be meaning analysis or structure analysis. It is not always meaning ambiguity. For example the word *jalan* has two categories, as a noun and a verb.

(1) *Sayamenemukandompét di jalan.* (Noun)

(2) *Jalan, yuk!* (Verb)

Ambiguity happens when words or phrases have more than one meaning (Fromkin 2014:142). Meanwhile, according to O'Grady (1997), several sentences are structurally ambiguous because the meanings of their component can be combined in several ways.

Sentence has influence on ambiguity. If an ambiguous word stands by itself, such as *jalan*, it can be ambiguous and cause many interpretations. But, if it is put in a sentence, it may not be ambiguous.

In this case, I will focus on some ambiguous prefixes. Generally, prefix does not have lexical meaning. It belongs to bound morpheme that cannot stand by itself (Carstairs 2002:18). For example, prefix *en-* does not have meaning if it stands alone, but after it is attached to base *large* and forming word *enlarge*, prefix *en-* gives new classes and meaning.

In Indonesian, there are ambiguous prefixes and non-ambiguous prefixes, as well. For example prefix *di-* in Indonesia can function as free morpheme, *kata depan* (preposition) and bound morpheme, *kata awalan* (prefix). *Kata depan* is written separately from the stem, while *kata awalan* is written connected with the stem.

- (1) **Ibumembelibajudipasar*
- (2) *Ibumembelibaju di pasar.*
- (3) *Baju ini dibeli oleh ibu.*
- (4) **Baju ini di beli oleh ibu*

2.3. Morphophonemic

2.3.1 Morpheme, Morphs, Allomorphs

Morpheme is the smallest meaningful linguistic unit. According to Katamba (1993), morpheme is the smallest unit of meaning. A morph is a physical shape that represents some morpheme in a language.

Ramelan (1992:121) wrote, “allomorphs are conditioned variants of a morpheme, the distribution of which is determined by environment” or in other hand allomorph is the phonetic and orthographical variations that represent the same morpheme. For example, past tense marker in English /-ed/ has three variations [-Id], [-t], and [-d]. The variations are conditioned by the final sound of the base.

In Indonesian, prefix *meN-* (N means nasal) has several variations: Ø, *meny-*, *meng-*, *mem-*, and *men-*. The changes are conditioned by the initial sound

of the base. For example, we will use \emptyset variation if it attaches to base with initial *r* (*merasa*) or *l* (*melakukan*).

2.3.2 Morphophonemic Process

Morphophonemic is the study about phonetic representation of morphemes in different contexts (Ramelan 1992:121). According to Verhaar (2012), in morphophonemic, phonemes determine the rules. Parera (1994) called morphophonemic as alternation. Parera in his book said that there are some conditions that determine why morphophonemic phenomena occurs. They are:

- (1) A morpheme is not always limited by one phonemic construction. A morpheme in a certain environment is represented and stated with other phoneme constructions.
- (2) Therefore, it needs to create a technical terms to differentiate the different phonemes from the same morpheme.
- (3) The preference of phonemic construction over its variations must be contextually determined.

Thus, morphophonemic is the study of morpheme and phoneme. There are also some steps in analysing the morphophonemic process :

- (1) Determining the environment
- (2) Looking for the similarities
- (3) Identifying the morpheme
- (4) Writing the rules

2.4 Machine Readable Linguistic Resource

Machine readable linguistic resource (MRLR) is a general term for resources which can be applied to annotate a corpus data. For example, BNC (British National Corpus), COCA (Corpus of Contemporary American English) and ICNALE annotated their own corpus by their MRLR and loaded the corpus data on their websites. However, the language of all those MRLRs is English, not Indonesian.

Most of the MRLR have default language which is English and users do not have access to modify the resources. So far resources in Indonesia are very limited and mostly built by computer scientists, such as Morphind (Larasati, 2011), iPosTagger (Farizky&Ayu, 2010) and PosTagID (Dinakaramani, 2014).

However, basic programming skills are required to exploit these resources. This is a problem for linguists who do not have any programming skill. In NooJ Application, we can write our own language MRLR without any programming skills. Grammars and dictionaries are used to create Indonesian lexical resources. Figure 3.1 and figure 3.2 show how NooJ handles inflection for English verbs. Same principle can apply to Indonesian with different method that comply with Indonesian language structure.

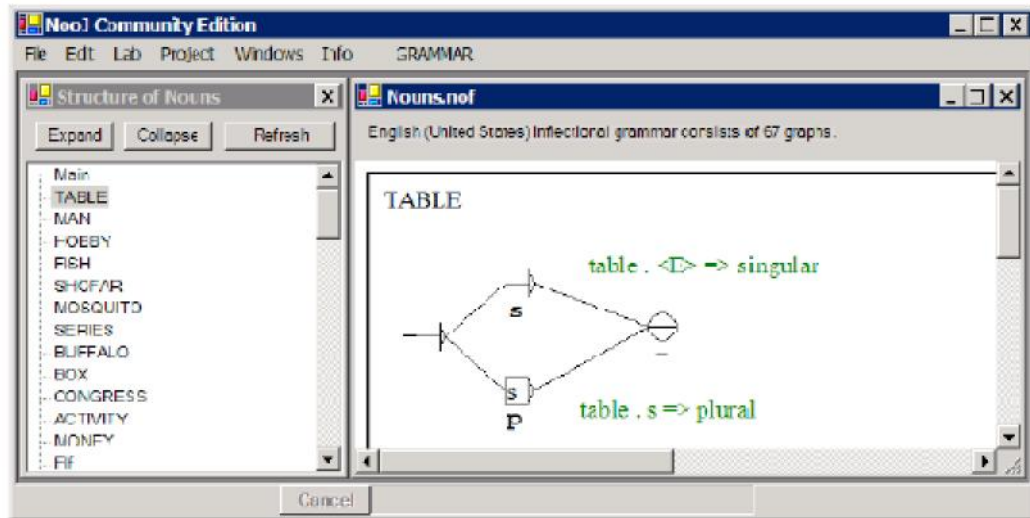


Figure 3.1 NooJ Grammar

`artist, N+FLX=TABLE+Hum`
`cousin, N+FLX=TABLE+Hum`
`pen, N+FLX=TABLE+Conc`
`table, N+FLX=TABLE+Conc`
`man, N+FLX=MAN+Hum`

Figure 3.2 NooJ Dictionary

CHAPTER III

RESEARCH METHODOLOGY

In this part, I will present the methods of research that I have done in conducting this research. This chapter includes corpus data, procedures and tags.

3.1 Data

The data that I used are journals where the papers are written in Indonesian. First, I identified the journal portals whether they are accredited by Dikti or not. Dikti stands for *Directorat Pendidikan Tinggi* which provides us the information about higher education in Indonesia. After that, I selected the accredited journals that would be used as my data because they are more qualified and scientific. I used 25 linguistic papers written in Indonesian which were taken from several journals, such as *Masyarakat Linguistic Indonesia (MLI)*, *LITERA*, *Linguistika*, *Humaniora*, *Bahasadan Sastra Indonesia*, and *Lingua Cultura* from 2012 up to 2016. All those papers were used as my corpus data. All of the journal papers are in .pdf format. These papers are the raw data of my research.

In collecting the raw data, I have converted the text in <.pdf> format to the <.txt> format by copying and pasting them to the notepad. Then I combined the texts in the journals that have been copied and saved it as .txt. As an illustration, look at the figure 3.1.

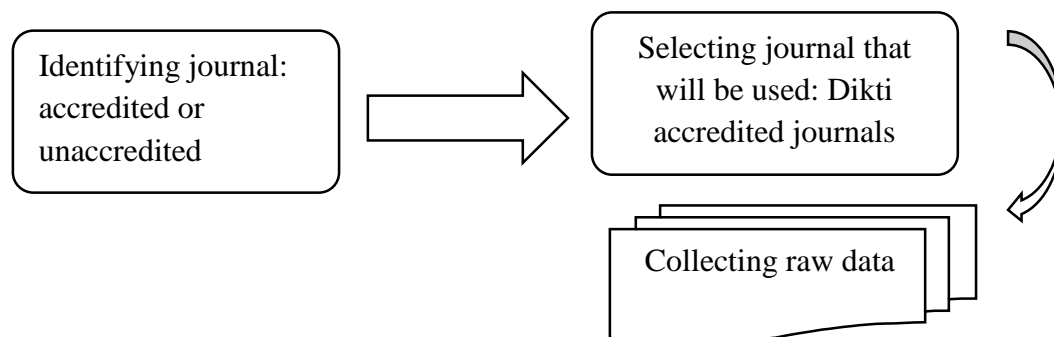


Figure 3.1 Process of collecting corpus data

3.2 Procedures

There were several procedures in conducting this research after the raw data was collected. They were collecting wordlist, reviewing literature, and analysing data.

3.2.1 Collecting Wordlist

The first was collecting word list. Wordlist is used as the resources for the annotation. In collecting the word list, firstly, I decided to use the word lists from *Kamus Besar Bahasa Indonesia (KBBI)*. I obtained all Indonesian nouns. Then, I took purposive sampling technique introspectively, which focus on frequently used nouns. There were 2048 frequently used nouns. After collecting the word list, I decided to choose what part of speech which became the result of the derivational process which is verb. Then, I selected prefixes which are used in changing nouns to verbs. To make this research specific, I focused on four verbalizer prefixes which are *meN*, *beR-*, *teR-* and *di-*.

3.2.2 Reviewing Literature

In this section, I reviewed several theories and references to support the data analysis. Reviewing some literatures has shown that some prefixes tend to be ambiguous. I also discussed the complexities of prefixes and machine readable language resource. Then, the product of reviewing the literature was tags that will be explained more details in the next chapter.

3.2.3 Analysing Data

After collecting wordlist and reviewing literature, the next was identifying derivational process with the help of Ms. Excel. In this stages, I was aided with Excel's binary matrices. After that, I wrote the grammatical and semantic formulation, which could be performed with the prefixes. The detailed process will be presented in chapter IV, the data analysis.

The data that has been analysed were presented in two ways: formal and informal. Formal presenting method is a method which presents the data analysis in the forms of sign and symbol. Meanwhile, informal presenting method is a method which presents the data analysis in forms of words only (Sudaryanto 1993:145).

3.3 Tags

The purpose of giving tags is to accommodate the users as wide as possible to retrieve text orthographically, by its part of speech and by the morphology.

The first is tags for part of speech. It is used to identify the result of the derivation and the base form of the part of speech. As this research explains the

derivation from noun to verb, so, the symbols that are used are N that means noun as the base and V that means verb as the result of derivation.

The second one is tags for the prefixes, including morphophonemic. It is used to explain the meaning of symbols in the morphophonemic process. For example, MENG+1 and MENG1 are differently written, so they have different meaning and function, as well. MENG+1 has symbol (+), which shows that it is used to retrieve semantic information. Meanwhile, MENG1 is used to retrieve derivation.