# Using Corpus and Dependency Tool to Investigate Degree Adverb in Indonesian

**Khairani Fajrianisa[1], Totok Suhardijanto[2]**
{khairani.fajrianisa71@ui.ac.id[1], totok.suhardijanto@ui.ac.id[2]}

[1,2] Universitas Indonesia, Jakarta, Indonesia

**Abstract.** Technology development these days also gives a contribution to linguistic research. The technology makes linguist easier in collecting data for research. One of the developing linguistic research methods is a corpus-based study. This study uses a corpus linguistics tool to help in collecting and analyzing data. This research focuses on the adverb of degree in the Indonesian language and aims to investigate it syntactically. The adverb which uses in this research, is *sangat* as this adverb is used frequently in the Indonesian language. IndonesianWaC corpus from Sketch Engine (http://www.sketchengine.eu) is used to collect data and to see the words that are modified by *sangat*. *Sangat* is well-known as a modifier of an adjective, but in reality, *sangat* also modifies a verb in the verbal phrase. Besides, this research attempts to analyze how the adverb of degree is generated in the syntactic tree by using the UDPipe tool as *sangat* is found modifying clause other than a phrase.

**Keywords:** *Adjective, Corpus, Dependency, Qualitative Adverb, Verb*

## 1. INTRODUCTION

Adverb of degree has a function as a degree modifier in a phrase. Degree modifiers can be defined as elements that modify another aspect to the degree [1]. In Indonesian, the adverb of degree is better known as a qualitative adverb. This adverb is used to describe meaning related to scale, degree, and quality [2]. Adverb of a degree in Indonesian is always considered as a modifier for an adjective. This adverb, especially *sangat* 'very', even is used as a parameter to distinguish verb and adjective. If an investigated word is modified by *sangat* and the phrase is grammatical, the word most likely is an adjective, e.g., *cantik* 'beautiful' will form grammatical phrase when it is modified by *sangat* as in phrase *sangat cantik* 'very beautiful'. Meanwhile, *lari* 'run' creates ungrammatical form when it is modified by *sangat* 'very' as in phrase *sangat lari* 'very run'.

As stated above, the adverb of degree is well known as a modifier for adjectives only. This opposite with the use of an adverb of a degree in the corpus. Adverb of degree is found modifying another part of speech, other than adjective as seen below.

(a)  Pasarnya sangat kecil. (doc#271)
*Market-the  very  small*
*The market is very small*

(b)    Bank syariah *sangat* diperlukan. (doc#20)
       Bank Islamic very    needed
       Islamic bank is very needed

In examples (a), it is expected that the adverb of degree, *sangat* 'very', is modified kecil 'small', which is an adjective. The combination forms a grammatical sentence. In example (b), *sangat* 'very' is modified passive verb diperlukan 'is needed', but it still forms grammatical construction, and its meaning even clearly can be understood. This fact contradicts a theory that says adverb of degree can be used to modify adjectives only. In Indonesian, this fact has been recognized by some linguists. Kridalaksana [3], Alwi [2], and Sneddon [4] found that there are verbs that can be modified by adverb of degree. They conclude that the verbs that can be modified by adverb of degree are the verb of emotion. Those verbs are emotion verbs only. It also contradicts with what is found in the corpus as *diperlukan* 'is needed' is not an emotion verb or causative verb formed by suffixation with the suffix *-kan*. Their conclusion contradicts the example above as a verb which is found modified by an adverb of degree above is not an emotion verb or causative verb with the suffix -kan. Thus, this research aims to explain the uses adverb of degree in the sentence, which included kind part of speech and the grammatical unit that is modified by the adverb. Corpus and dependency tool is used to investigate adverb of degree behavior.

## 2. METHOD

This research is corpus-based grammar research. Corpus-based grammar research uses the area that has been neglected by traditional studies. Instead of using institution or native judgment, corpus-based approach analyses the patterned use of grammatical features in texts. The availability of large corpora and computer tools make it possible to study the patterned ways in which speakers use the grammatical resources of a language – by investigating the frequency distribution of various constructions, the association patterns between grammatical structures and other linguistic and non-linguistics factors, and the factors that affect choices between structural variants [5]. By using large corpora, getting various patterns which are usually used by language speaker is possible so that the grammar description will be more precise than using native knowledge which must have some limitations.

There are several corpora which provide Indonesian corpus for Indonesian grammar research. Corpora, which has large Indonesian corpora, are Leipzig Corpora (https://corpora.uni-leipzig.de) and Sketch Engine (https://www.sketchengine.eu/). Leipzig corpora just provide use occurrences without generating collocation or frequency. Sketch Engine, which is used for this research, has more features than Leipzig. Sketch Engine provides us part of speech annotation, frequency, and list of collocations. The features make it easier for grammarians to do grammar analysis. Indonesian corpora in Sketch Engine is named Indonesian Web (Indonesian WaC). This corpus has 109,236,814 tokens generated from 27,051 internet articles. The internet articles vary, ranging from Wikipedia articles, blogs, to online news. This is one of the strengths of the corpus that provides data from various registers. As this research concerns the head of degree modifier *sangat*, the collocation feature is used. Degree modifier, in this case, is *sangat*, is always placed in front of the head or left of the head as follows.

(c) Sangat/modifier pintar/head

Very          clever

The right collocation of *sangat* will be used for this research. The range of collocation used for this research is one that means the generated collocation is placed right beside *sangat*.



Fig. 1. Setting to generate collocation of *sangat*



Fig. 2. Collocation list of *sangat* which is generated by Sketch Engine

The adverb of degree will be analyzed syntactically by using the dependency tool. Corpus will provide us with data, pattern, frequency, and collocation, while the dependency tool will help grammarian to analyze syntactic relation between constituent. One of the dependency tools which already has data for Indonesian is UDPipe (http://lindat.mff.cuni.cz/services/udpipe/info.php). UDPipe is a trainable pipeline for tokenization, tagging, lemmatization, and dependency parsing. This tool will generate part of the speech and dependency relation of construction in the form of the dependency tree.
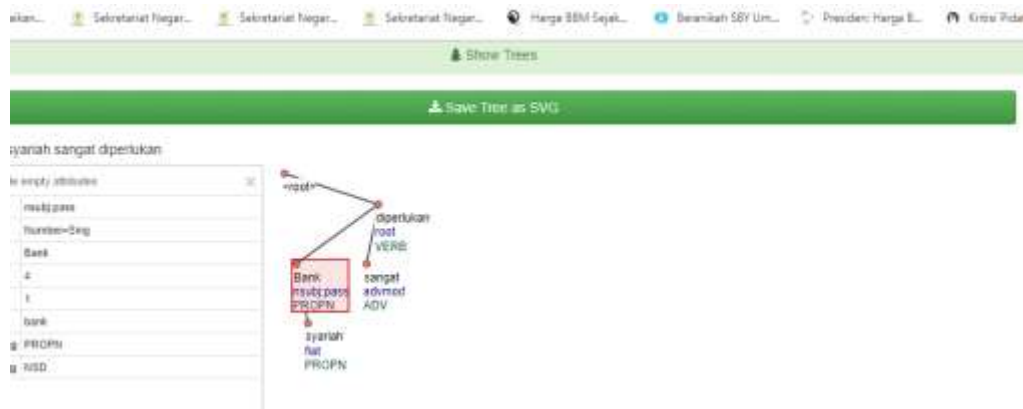
Fig.3. Dependency tree in UDPipe

## 3. RESULT AND DISCUSSION

Descriptive grammarians tend to use "field methods" to identify the various paradigms in language, while theoretical grammarians typically rely on their institutions about language and sometimes supplemented by asking native speakers to judge if made-up sentences are grammatical or not [5]. It may cause the grammar description has a limitation as to what happens with the characteristic of adverb of degree in Indonesian. Thus, deviation from the norm is generally considered to be errors by prescriptive approaches, or the deviation will be treated as the existence of grammatically correct alternatives by descriptive approaches, but often is given little attention to functional reasons for choosing between alternatives [5]. What happened to adverb of degree in Indonesian is caused by this descriptive approach since this deviation has been recognized by some linguists, but there is no complete explanation.

As stated above, *sangat* 'very' as an adverb of degree is always considered as a modifier of an adjective. However, in the corpus, it is found the adverb is collocated with another part of speech (PoS) than adjective. This table shows PoS of the collocation of *sangat* and its total occurrences in the corpus.

**Table 1.** PoS of collocation and total occurrences

| PoS | Amount | Occurrences |
|---|---|---|
| Pronoun | 7 | 294 |
| Adjective | 542 | 72217 |
| Verb | 411 | 46447 |
| Adv | 17 | 2954 |
| Misc. | 23 | 684 |
| | 1000 | 122596 |

Collocates, which is found right beside *sangat* are pronoun, adjective, verb, and adverb. Miscellaneous (Misc.) is a category that is made to mark punctuation and Malaysian language. In Sketch Engine, there are several Malaysian articles, so that manual review is needed to avoid
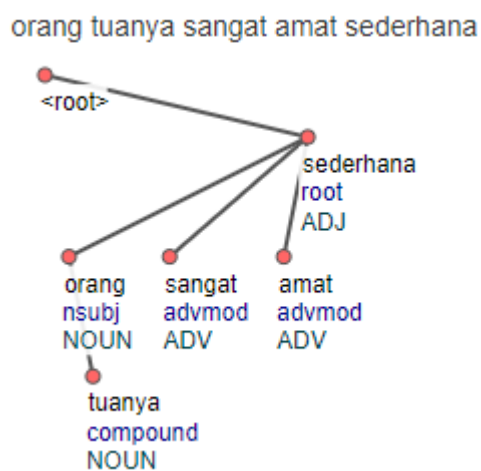
Malaysian word included in data. The tool can not distinguish punctuation as non-word, so punctuations are still generated as one of the collocates. As expected, an adjective is the highest PoS, which occurs with *sangat* followed by a verb. Klein is found that adjectives, verbs, and nouns are the main categories with the gradable feature. Noun is not found as a collocate of *sangat*. Thus it can be concluded that a noun is PoS that does not accept degree modifying.

There is two part of speeches which are different from what has been found by Klein [6]. Pronoun and adverb are found as collocates of *sangat* in data. The frequency of pronoun and adverb is not as high as the frequency of adjectives and verbs. Below are examples of sentences with pronoun and adverb collocated with *sangat*.

(d) Orang tuanya ***sangat* amat** sederhana (doc#112)

parents-his     very   very   humble

parents-his is very humble

(e) Chairul adalah orang yang **sangat kami** butuhkan (doc#474)

Chairul is     someone  that very we   need

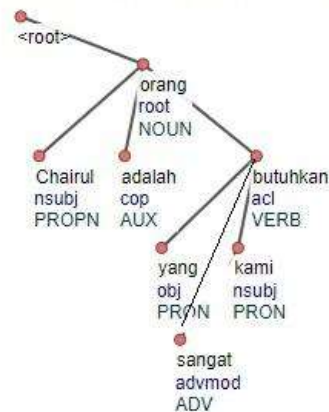Chairul is someone that we really need

Example [d] is an example of two adverbs in one sentence. Although *amat* appears as the collocation of *sangat* in the corpus, it does not mean *sangat* modifies *amat* and form phrase. *Sangat amat* is an example of two adverbs in one sentence in Indonesian. Both *sangat* and *amat* modify adjective *sederhana* as in this syntactic structure.



**Fig.4. Dependency of *sangat amat***

In example (d), *sangat* does not modify *kami*, but modify the verb *butuhkan* 'need' as the sentence can be paraphrased as *Chairul adalah orang yang kami sangat butuhkan.* Thus, *sangat* is still considered as a modifier of *butuhkan* 'need' instead of *sangat*.

**Fig.5. Dependency of *sangat* and *kami***

Dependency trees above also can be used to see the realization of the adverb position in a sentence. Greenbaum, et al. [7] found that adverb can realize in the initial position, medial position, in split infinitive construction, and end position. Adverb of degree in Indonesian only occurs in medial position. Medial position can be preliminarily described as adverb which is between S and V. Medial posistion is mainly associated with modality and degree, such that adverbials tend to take on a tinge of those even when they purport to be expressing other relations [7]. Adverb of degree in Indonesian is found in between subject and verb as seen in examples below.

> (f) Bank syariah sangat diperlukan… (doc#20)
>     [lit.] Islamic bank is very needed

> (g) … cara ini sangat membantu Anda untuk menjalankan diet (doc#186)
>     [lit.] This technique very help you to do diet

From the examples above, *sangat* 'very' always occurs before the verb which is its modified. In example (f), *sangat* 'very' occurs before passive verb *dibutuhkan* 'needed'. In example (g), adverb of degree *sangat* 'very' is found before verb *membantu* 'help'. Medial

## 4. CONCLUSION

According to the collocate of *sangat*, it is found that there is a category that can be modified by *sangat*. The categories are adjective and verb. Although there are other categories which are found in *sangat*, those categories are not proven as head of degree modifier based of its dependency relation. Adverb of degree, *sangat*, is found as a modifier of a word only in phrase construction. This adverb is not found as a clause or sentence modifier. In sentence structure, it is found that this adverb always takes a position in medial, between subject and verb.

# REFERENCE

[1]     C. Paradis, *Degree modifiers of adjectives in spoken British English*. Lund University Press, 1997.

[2]     H. Alwi, S. Dardjowidjojo, H. Lapoliwa, and A. M. Moeliono, *Tata Bahasa Baku Bahasa Indonesia (Indonesian Grammar)*, Ketiga. Jakarta: Balai Pustaka, 2003.

[3]     H. Kridalaksana, *Kelas kata dalam bahasa Indonesia*. Jakarta: PT Gramedia Utama, 2007.

[4]     J. N. Sneddon, A. Adelaar, D. N. Djenar, and M. C. Ewing, *Indonesian reference grammar 2nd edition*. Sydney: Allen & Unwin, 2010.

[5]     K. Precht, D. Biber, S. Conrad, and R. Reppen, *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press, 1998.

[6]     H. Klein, *Adverbs of Degree in Dutch and Related Languages*. J. Benjamins Pub., 1998.

[7]     R. Huddleston, R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik, *A Comprehensive Grammar of the English Language*. London: Longman, 1985.